Copulae and tail dependence

Diploma thesis

in partial fulfillment of the requirements
for the degree of **Diplom-Volkswirt**

by

Till Großmaß

Matriculation number: 186172

submitted to
Prof. Dr. Wolfgang Härdle
Institute for Statistics and Econometrics
School of Business and Economics
Humboldt-University Berlin

Berlin, September 28, 2007

Abstract

This thesis presents the concept of tail dependence in a financial context as one tool to measure dependence in the extremes of a bivariate distribution. Copulae can separate the problem of estimating a multidimensional distribution into the estimation of the marginal distributions and the dependence between the one-dimensional random variables. Therefore, copulae are used in order to carry out the estimation of the tail dependence coefficient (TDC). Four estimators of the TDC are presented and compared in a simulation study for various distributions and copulae. Furthermore, an introduction into bivariate Extreme Value Theory (EVT) is given, which tries precisely to analyze the behavior at the tail of a bivariate distribution. EVT allows to construct estimators of the TDC and to derive a test for tail independence, which is recognized to be indispensable but rarely utilized in a financial context. As an application to nine different financial data sets shows, the phenomenon of tail dependence is less common than often argued in the literature: the periods where indeed tail independence can be rejected are few.

Keywords:
copulae, tail dependence, bivariate extreme value theory, test for tail independence

Declaration of Authorship

I hereby confirm that I have authored this diploma thesis independently and without use of others than the indicated sources. All passages which are literally or in general matter taken out of publications or other sources are marked as such.

Berlin, September 28, 2007

Till Großmaß

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A Google search on copula yields in september 2007 approximately 1,310,000 results whereas there were 10,000 results in 2003 according to Mikosch [2005]. This increase comes from the mathematical (or statistical) meaning of the word (not from a copula boom in linguistics where the word copula is also used). In the literature as well, many recent financial applications of the theory of copulae and tail dependence can be found, e.g. Fernandez [2005], Junker and May [2005], Kole et al. [2007] or Rodriguez [2007]. Furthermore, the recent sub-prime mortgage crisis shows that risk management is an important sector where further research can help to avoid global crashes. One tool in this context can be tail dependence, in order to assess the probability of unlikely events that occur jointly.

Copulae are able to model multivariate distributions in an easy way and their use is more and more widespread in different applications, among others in the financial sector. Copulae can separate the problem of estimating a multidimensional distribution into the estimation of the marginal distributions and the dependence between the one-dimensional random variables. This can be useful when one is mainly interested in the dependence between different assets, as for example in risk management. In this case, it is important to measure the extent to which different assets are linked and of course different ways to do so exist. The concept presented in this thesis is tail dependence, which describes the limiting probability that one random variable exceeds a certain threshold given that another random variable has already exceeded that specific threshold.

The structure of this thesis is as follows. The second chapter introduces the concept of copulae, some examples of copulae are given and the way a copula model can be estimated is explained. In the third chapter, the concept of tail dependence is introduced and linked to copulae. Different ways of estimating the tail dependence coefficient (TDC) are presented, mainly based on bivariate Extreme Value Theory. Afterwards, a test for tail independence is introduced, which is indispensable when working with tail dependence, since all estimators of the TDC are strongly misleading when the data does not stem from a tail dependent setting. In chapter 4, the methods are checked with simulated data and in chapter 5 they are applied to different financial assets. The thesis ends with a conclusion in chapter 6.

Chapter 2

Copulae

This chapter introduces the concept of copula, which will prove to be helpful for the analysis of tail dependence. First, the definition of a copula is given and the most important theorem concerning copulae is stated: Sklar's theorem, which illustrates furthermore the origin of the word copula: the fact that a copula couples marginal distributions with the joint distribution. Afterwards, two important classes of copulae are presented, Archimedean and elliptically contoured copulae (which include Gaussian copula and t-copula). As shown later, these two types are able to model upper and lower tail dependence, but only in one direction (e.g. Gumbel or Clayton) or symmetrically (e.g. t-copula or Gaussian copula). Therefore, mixture copulae are introduced as a means to model asymmetric tail dependence, which could be expected to be found in a financial context: large joint losses of two assets are intuitively more often than large joint gains. In the following section, different techniques to estimate the parameters of copulae are presented as well as the way to simulate random variates from copulae. In section 2.5 a description of the problem of how to choose one copula model among different alternatives is given. The chapter ends with an example of the copula estimation for switching parameters.

2.1 Basic definitions and Sklar's Theorem

A copula is a multivariate distribution function with uniform margins. The formal definition is the following one.

Definition 2.1.1. *A function $C : [0,1]^d \to [0,1]$ is called a d-dimensional copula if and only if:*

1. $\forall u = (u_1, \dots, u_d)^\top \in [0,1]^d$, $C(u_1, \dots, u_d) = 0$ *if* $u_j = 0$ *for at least one* $j \in \{1, \dots, d\}$ *(C is grounded).*

2. $\forall u = (u_1, \dots, u_d)^\top, \forall v = (v_1, \dots, v_d)^\top \in [0,1]^d$, *such that* $\forall j \in \{1, \dots, d\}, v_j \leq u_j$,

$$V_C(B) := D_{u_d}^{v_d} \dots D_{u_1}^{v_1} C(t_1, \dots, t_d) \geq 0$$

where $B = ([u_1, v_1] \times \cdots \times [u_d, v_d])$, $t \in [0,1]^d$ *and* $D_{u_j}^{v_j} C(t_1, \ldots, t_d) = C(t_1, \ldots, t_{j-1}, v_j, t_{j+1}, \ldots, t_d) - C(t_1, \ldots, t_{j-1}, u_j, t_{j+1}, \ldots, t_d)$ *(C is d-increasing).*

3. *C has margins* C_k *satisfying* $\forall k = 1, \ldots, d$*:* $C_k(u) = u, \forall u \in [0,1]$.

For studying tail dependence, mostly bivariate distributions are used. Therefore the copula definition is also given for the bivariate case:

Definition 2.1.2. *A function* $C : [0,1]^2 \to [0,1]$ *is called a 2-dimensional copula if and only if*

1. $\forall u \in [0,1], C(0,u) = C(u,0) = 0.$
2. $\forall u \in [0,1], C(u,1) = C(1,u) = u.$
3. $\forall (u_1, u_2), (v_1, v_2) \in [0,1]^2$ *with* $u_1 \leq v_1$ *and* $u_2 \leq v_2$ *:*

$$C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0.$$

The concept of copulae being defined, we can now come to the eponymous copula theorem from Sklar [1959].

Theorem 2.1.1. *(Sklar) Let F be a d-dimensional distribution function with margins* $F_1, \ldots, F_d$*. Then there exists a d-dimensional copula C such that* $\forall x \in \bar{\mathbb{R}}$*:*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

If all the margins are continuous then C is unique. Conversely, if C is a d-dimensional copula and $F_1, \ldots, F_d$ *are distribution functions, then F is a distribution function with margins* $F_1, \ldots, F_d$*.*

Proof. See Sklar [1959]. □

In order to formulate the likelihood function of the estimation problem it is indispensable to define the copula density, which can be done as follows:

Definition 2.1.3. *If* $\forall (u_1, \ldots, u_d)^\top \in [0,1]^d$,

$$C(u_1, \ldots, u_d) = \int_0^{u_1} \cdots \int_0^{u_d} \frac{\partial^d C(t_1, \ldots, u_t; \theta)}{\partial t_1 \ldots \partial t_d} dt_1 \ldots dt_d$$

C is said to be absolutely continuous and we can define the copula density c by, $\forall (u_1, \ldots, u_d)^\top \in]0,1[^d$*:*

$$c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}.$$

A helpful proposition for copulae is the following invariance property.

Proposition 2.1.1. *A copula C is invariant under strictly, monotone increasing transformation.*

Proof. See Nelsen [2006], p. 25. □

This implies that for example, price and log-price of a specific financial asset have the same copula. To simplify notation, the survival copula is defined as follows (in the same way as for distribution functions in general):

Definition 2.1.4. *The survival copula $\tilde{C}$ is defined as:*

$$\tilde{C}(u,v) = u + v - 1 + C(1-u, 1-v).$$

2.2 Different types of copulae

Of course, many different classes of copulae exist. Here, two important ones are defined: Archimedean and elliptically contoured copulae. Archimedean copulae have the advantage that they are described by only one parameter. This simplicity comes of course at a cost: they are less flexible.

2.2.1 Archimedean copulae

In order to define Archimedean copulae, the notion of pseudo-inverse is to be given:

Definition 2.2.1. *Let φ be a continuous, strictly decreasing function from $[0,1]$ to $[0,\infty]$ such that $\varphi(1) = 0$. $\varphi^{[-1]}$ is called pseudo-inverse of φ and given by, $\forall t \in [0,\infty]$:*

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \le t \le \varphi(0), \\ 0, & \varphi(0) \le t \le \infty. \end{cases}$$

With this pseudo-inverse, Archimedean copulae can be defined as follows:

Definition 2.2.2. *A 2-dimensional copula C is called an Archimedean copula if and only if there exists a continuous, strictly decreasing, convex function φ (called generator) from $[0,1]$ to $[0,\infty]$ such that $\varphi(1) = 0$ and*

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2)).$$

If additionally $\varphi(0) = \infty$ and hence $\varphi^{[-1]} = \varphi^{-1}$, C is called strict Archimedean copula and φ strict generator.

Some examples of Archimedean copulae are given below:

Example 2.2.1. *The 2-dimensional Clayton copula is defined as follows, $\forall \theta > 0$:*

$$C^C(u_1, u_2; \theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\theta^{-1}}.$$

Example 2.2.2. *The 2-dimensional Frank copula is defined as follows, $\forall \theta > 0$:*

$$C^F(u_1, u_2; \theta) = -\frac{1}{\theta} \log\left[1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right].$$

Example 2.2.3. *The 2-dimensional Gumbel copula is defined as follows,* $\forall \theta \in [-1, \infty)$*:*

$$C^G(u_1, u_2; \theta) = \exp[-\{(-\log u_1)^\theta + (-\log u_2)^\theta\}^{\frac{1}{\theta}}].$$

2.2.2 Elliptically contoured copulae

Now, define a more complex class of copulae, which is widely used.

Definition 2.2.3. *Let X be an n-dimensional random vector, $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ a symmetric positive semi-definite matrix. If the characteristic function $\varphi_{X-\mu}(t)$ of $X - \mu$ is of the form $\varphi_{X-\mu}(t) = \Psi(t^\top \Sigma t)$ where $\Psi : \mathbb{R}^{+*} \to \mathbb{R}$, then X is said to have an elliptically contoured (or elliptical) distribution and we write $X \sim E_n(\mu, \Sigma, \Psi)$. Ψ is called characteristic generator.*

The following theorem (see Embrechts et al. [2001]) characterizes the class of elliptical distributions.

Theorem 2.2.1. *$X \sim E_n(\mu, \Sigma, \Psi)$ with $rank(\Sigma) = k$ if and only if there exists a random variable $R \geq 0$ independent of U, a k-dimensional random vector uniformly distributed on the unit hyper-sphere $\{z \in \mathbb{R} | z^\top z = 1\}$, and an $n \times k$ matrix A with $AA^\top = \Sigma$, such that:*

$$X \overset{d}{=} \mu + RAU.$$

Proof. See Embrechts et al. [2001]. □

Using definition 2.2.3 and the characterizing theorem 2.2.1 it can be seen (for a proof see again Embrechts et al. [2001]) that the Gaussian copula and the t-copula belong to the elliptically contoured copulae. Since these two copulae are widely used, they are given below.

Example 2.2.4. *The t-copula is defined as follows:*

$$C^t(u_1, u_2; \nu, \rho) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \times \left\{1 + \frac{s_1^2 - 2\rho s_1 s_2 + s_2^2}{\nu(1-\rho^2)}\right\}^{-(\nu+2)/2} ds_1 ds_2.$$

Example 2.2.5. *The Gaussian copula is given by*

$$C^N(u_1, u_2; \rho) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \times \exp\left\{-\frac{s_1^2 - 2\rho s_1 s_2 + s_2^2}{2(1-\rho^2)}\right\} ds_1 ds_2.$$

2.2.3 Copula mixtures

For modeling asymmetric tail dependence, Archimedean and elliptical copulae are not sufficient. Copula mixtures can account for asymmetries and can be easily defined.

Proposition 2.2.1. *A linear convex combination of 2 copulae is still a copula:*

$$C(u_1, \ldots, u_d; \theta) = \lambda C_1(u_1, \ldots, u_d; \theta_1) + (1 - \lambda) C_2(u_1, \ldots, u_d; \theta_2),$$

and analogously the density copula:

$$c(u_1, \ldots, u_d; \theta) = \lambda c_1(u_1, \ldots, u_d; \theta_1) + (1 - \lambda) c_2(u_1, \ldots, u_d; \theta_2).$$

Proof. See Nelsen [2006], p. 72-73. □

Therefore, mixture copulae can easily be constructed by combining two different copulae. Throughout this thesis, the mixture copulae from example 2.2.6 are used. Figure 2.1 gives examples of simulated samples of different copulae.

Example 2.2.6. *Mixture of Clayton and Gumbel:*

$$C^{CG}(u_1, u_2; \theta_1, \theta_2, \lambda) = \lambda C^{C}(u_1, u_2; \theta_1) + (1 - \lambda) C^{G}(u_1, u_2; \theta_2).$$

Mixture of Clayton and Survival Clayton:

$$C^{C\tilde{C}}(u_1, u_2; \theta_1, \theta_2, \lambda) = \lambda C^{C}(u_1, u_2; \theta_1) + (1 - \lambda) \tilde{C^{C}}(u_1, u_2; \theta_2).$$

Mixture of Survival Gumbel and Survival Clayton:

$$C^{\tilde{G}\tilde{C}}(u_1, u_2; \theta_1, \theta_2, \lambda) = \lambda C^{\tilde{G}}(u_1, u_2; \theta_1) + (1 - \lambda) C^{\tilde{C}}(u_1, u_2; \theta_2).$$

Mixture of Gumbel and Survival Gumbel:

$$C^{G\tilde{G}}(u_1, u_2; \theta_1, \theta_2, \lambda) = \lambda C^{G}(u_1, u_2; \theta_1) + (1 - \lambda) C^{\tilde{G}}(u_1, u_2; \theta_2).$$

2.3 Estimation

Sklar's theorem proves the existence of a d-dimensional copula C such that:

$$F(x_1, \ldots, x_d) = C(F_{X_1}(x_1), \ldots, F_{X_d}(x_d); \theta),$$

where F denotes the multivariate cumulative density function (cdf), F_{X_i} the marginal distribution of X_i and θ is the copula parameter (vector). The

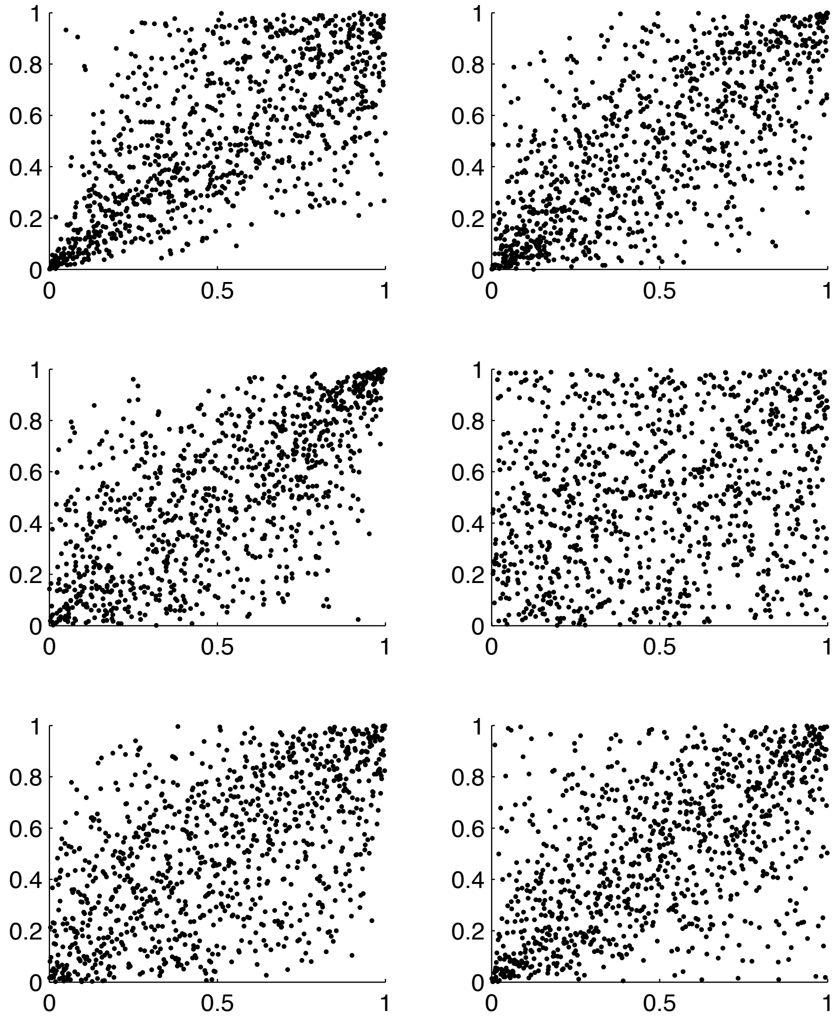

Figure 2.1: Simulated samples of size $1,000$ for Clayton ($\theta = 2$), Mixture of Clayton and Gumbel ($\theta_1 = 2, \theta_2 = 2, \lambda = 0.5$), Gumbel ($\theta = 2$) and Plackett ($\theta = 2$) copula and from Gaussian ($\rho = 0.6$) and t ($\rho = 0.6, \nu = 3$) distribution (transformed to $[0, 1]$ by transforming the margins to $[0, 1]$ by means of the empirical cdf); from left to right and top to bottom

density of the copula C is given by:

$$c(u_1, \ldots, u_d; \theta) = \frac{\partial^d C(u_1, \ldots, u_d; \theta)}{\partial u_1 \ldots \partial u_d}.$$

Then Sklar's theorem can be written in terms of density functions:

$$f(x_1, \ldots, x_d) = c\{(F_{X_1}(x_1), \ldots, F_{X_d}(x_d); \theta)\} \prod_{j=1}^{d} f_j(x_j).$$

The likelihood function for one observation $(x_{1,i}, \ldots, x_{d,i})$ is:

$$\begin{aligned} L_i(x_{1,i}, \ldots, x_{d,i}) &= f(x_{1,i}, \ldots, x_{d,i}) \\ &= c\{(F_{X_1}(x_{1,i}), \ldots, F_{X_d}(x_{d,i}); \theta)\} \prod_{j=1}^{d} f_j(x_{j,i}). \end{aligned}$$

The likelihood function for n observations:

$$L(x_1, \ldots, x_d) = \prod_{i=1}^{n} f(x_{1,i}, \ldots, x_{d,i}),$$

and therefore the log-likelihood function:

$$l(x_1, \ldots, x_d) = \sum_{i=1}^{n} \left[\log c\{(F_{X_1}(x_1), \ldots, F_{X_d}(x_d); \theta)\} + \sum_{j=1}^{d} \log f_j(x_j) \right].$$

Different algorithms to estimate the parameters exist: Full Maximum Likelihood, Inference for Margins and semiparametric approaches. In the next sections, some details are provided for these techniques. For a comparison of the different estimation techniques see e.g. Kim et al. [2007].

2.3.1 Full Maximum Likelihood

The log-likelihood is maximized to obtain the joint estimates of the parameters. In a parametric approach, one has to specify the marginal distributions of $X_1, \ldots, X_d$ to obtain $f_i, F_{X_i}, \forall i = 1, \ldots, d$. This parametric approach can be computationally intensive.

2.3.2 Inference for Margins

Another parametric method, which is faster than Full Maximum Likelihood is the Inference for Margins method (IFM). In a first step, the parameters of the marginal distributions are estimated via maximum likelihood, then maximum likelihood is carried out to estimate the copula parameter (given the estimated marginal distributions).

2.3.3 Semiparametric approach

For both methods (Full ML and IFM), the distribution of the margins has to be specified. A more flexible method consists of using a nonparametric estimator for the marginal distributions (the empirical distribution function) and then to estimate the copula parameters with maximum likelihood.

The estimator of the marginal distributions is the following empirical distribution function:

$$\tilde{F}_{nj}(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathrm{I}(X_{ji} \leq x), \forall j = 1, \ldots, d$$

and substituting this estimator in the log-likelihood function gives the maximization problem to estimate the copula parameters. This estimator is given by (for details, see Chen and Fan [2006]):

$$\tilde{\theta} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} \log c\{(\tilde{F}_{n1}(x_1), \ldots, \tilde{F}_d(x_d); \theta).$$

2.4 Monte Carlo Simulation of copulae

A general approach (see Embrechts et al. [2001]) to simulate variates from copulae uses the fact that a copula has uniform margins and that a distribution function is uniformly distributed. First, define the k-dimensional margin of a copula.

Definition 2.4.1. *The k-dimensional margin of a d-dimensional copula is given by*

$$C_k(u_1, \ldots, u_k) = C(u_1, \ldots, u_k, 1, \ldots, 1), \forall k \in \{2, \ldots, d-1\},$$

and $C_1(u_1) = u_1, C_d(u_1, \ldots, u_d) = C(u_1, \ldots, u_d)$.

Now, the conditional distribution of U_k given $U_1 = u_1, \ldots, U_{k-1} = u_{k-1}$ is:

$$\begin{aligned} C_k(u_k|u_1, \ldots, u_{k-1}) &= \mathrm{P}(U_k \leq u_k | U_1 = u_1, \ldots U_{k-1} = u_{k-1}) \\ &= \frac{\partial^{k-1} C_k(u_1, \ldots, u_k)}{\partial u_1 \ldots \partial u_{k-1}} \left\{ \frac{\partial^{k-1} C_{k-1}(u_1, \ldots, u_{k-1})}{\partial u_1 \ldots \partial u_{k-1}} \right\}^{-1}. \end{aligned}$$

The algorithm consists then of generating U_1 from a uniform distribution $\mathrm{U}[0,1]$ then U_2 from $C_2(.|u_1)$ etc. and U_d from $C_d(.|u_1, \ldots, u_{d-1})$.

The following two examples show how this algorithm can be implemented in the case of a Clayton copula and for a mixture of Clayton and Gumbel copulae.

Example 2.4.1. *If we want to simulate a bivariate Clayton distributed variable,* *we first have to calculate the conditional distribution of* $U_2|U_1 = u_1$*:*

$$\begin{aligned} C_2(u_2|u_1) &= \frac{\partial C_2(u_1,u_2)}{\partial u_1}\left\{\frac{\partial C_1(u_1)}{\partial u_1}\right\}^{-1} = \frac{\partial C_2(u_1,u_2)}{\partial u_1} \\ &= \frac{\partial}{\partial u_1}(u_1^{-\theta}+u_2^{-\theta}-1)^{-\theta^{-1}} = u_1^{-\theta-1}(u_1^{-\theta}+u_2^{-\theta}-1)^{-\theta^{-1}-1} \\ &= \{u_1^{\theta}(u_1^{-\theta}+u_2^{-\theta}-1)\}^{-\frac{1+\theta}{\theta}} = \{1+u_1^{\theta}(u_2^{-\theta}-1)\}^{-\frac{1+\theta}{\theta}}. \end{aligned}$$

Solving the equation $q = C_2(u_2|u_1)$ *for* u_2 *yields:*

$$u_2 = \left\{1 - u_1^{-\theta} + (u_1 q^{\frac{1}{1+\theta}})^{-\theta}\right\}^{-\frac{1}{\theta}}.$$

Now, q *and* u_1 *are drawn from a uniform distribution,* c_2 *is chosen according to the above formula, which yields a vector* (U_1, U_2) *of Clayton distributed variables with dependence parameter* θ*.*

Example 2.4.2. *In the case of a simple mixture of Clayton and Gumbel, we already see that the general algorithm is not always appropriate. First, calculating the following conditional distribution is not trivial.*

$$\begin{aligned} \frac{\partial C_2^{CG}(u_1,u_2)}{\partial u_1} &= \lambda\{u_1^{-\theta_1-1}(u_1^{-\theta_1}+u_2^{-\theta_1}-1)^{-\theta_1^{-1}-1}\} \\ &\quad + (1-\lambda)\left\{\exp[-\{(-\log u_1)^{\theta_2}+(-\log u_2)^{\theta_2}\}^{\frac{1}{\theta_2}}]\right. \\ &\quad \times \left.\left[\{(-\log u_1)^{\theta_2}+(-\log u_2)^{\theta_2}\}^{\theta_2^{-1}-1}(-\log u_1)^{\theta_2-1}\frac{1}{u_1}\right]\right\} = q. \end{aligned}$$

Second, solving this equation for u_2 *is not possible analytically. A numerical minimization algorithm is in this simple case possible, but can be burdensome in more advanced problems.*

These two examples show that a general algorithm is not always the best. Embrechts et al. [2001] give the algorithm for simulating from a t-copula and the Gaussian copula. The simulation from a bivariate Archimedean copula can be done using the general algorithm above. For mixture copulae, simulation is straightforward as well. Suppose we know how to simulate from copula C_1 and C_2. Then for given mixture parameter λ, we simulate from C_1 if $q \leq \lambda$ and from C_2 else, where Q is drawn from a uniform distribution on $[0, 1]$.

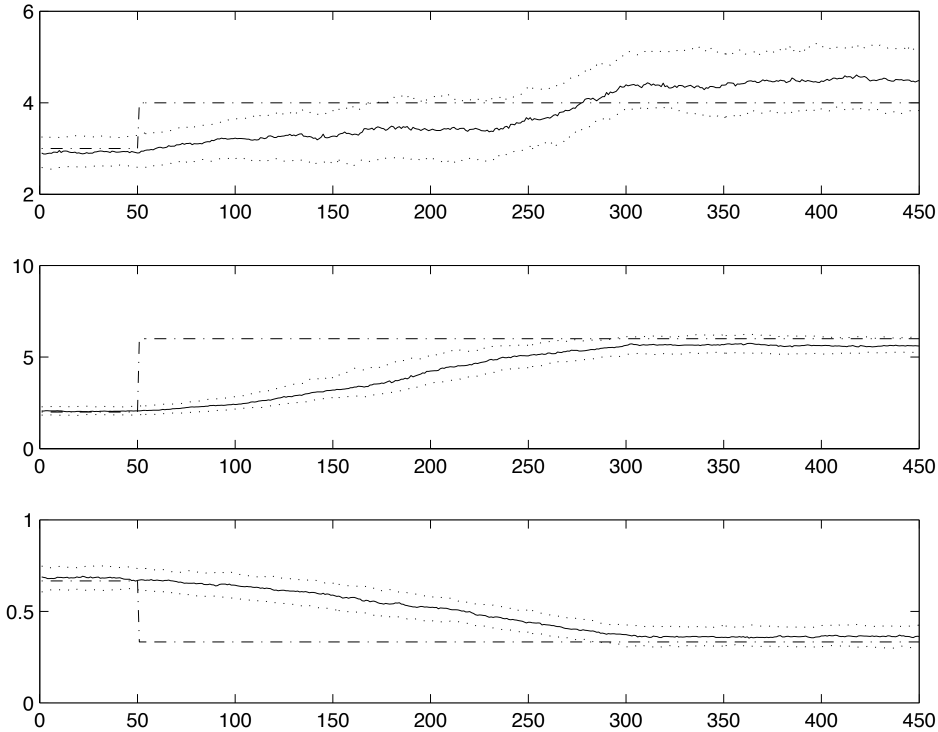

Figure 2.2: Mixture of Clayton and Gumbel copula. 200 simulations are made and median (full line), 25%- quantile and 75%- quantile (dashed) are calculated. The true parameters are dash-dotted.

2.5 Model selection

A difficulty that arises when working with copula models is the question of how to select the model that best fits the data. To my knowledge, no statistical criterion has yet been developed in this specific context. Nevertheless, Dias and Embrechts [2004] and Chen and Fan [2006] use the Akaike Information criterion (AIC), which is defined as follows:

Definition 2.5.1. *The Akaike Information Criterion is given by*

$$AIC = -2l(x_1, x_2; \hat{\theta}) + 2q,$$

where

$$l(x_1, x_2; \theta) = \sum_{i=1}^{n} \Big[\log c\{(F_{X_1}(x_1), F_{X_2}(x_2); \theta)\} \Big],$$

is the pseudo log-likelihood and q is the number of parameters. Lower values of the AIC indicate a better model.

According to the AIC, the model with the lowest AIC is chosen. Chen and Fan [2006] propose a pseudo likelihood ratio test to test if the difference in the AIC is significant or not. The problem of the AIC is that it is developed to test within nested models. There is no theoretical reason why likelihoods can be compared as done by the AIC in order to assess the quality of one model or another. Furthermore, since this thesis focuses more on tail dependence, the AIC is not used.

2.6 Fitting a copula model for simulated data

Figure 2.2 presents the results for a mixture of a Clayton and a Gumbel copula. The parameters switch from $\theta_1 = 3, \theta_2 = 2, \lambda = 2/3$ to $\theta_1 = 4, \theta_2 = 6, \lambda = 1/3$. The procedure is checked for a window length of 250, the first sample consists of 300, the second one of 400 observations, i.e. the parameters jump at $T = 300$. 200 simulations are made and median, 25%- quantile and 75%- quantile are calculated. It can be seen that the detection delay of the procedure is of about one window length.

Until now, we introduced the basic concepts of copulae and how a copula model can be estimated. In the next chapter we will use the copula concept to deal with tail dependence.

Chapter 3

Tail dependence

This chapter introduces the concept of tail dependence, which is important in modeling dependence of extreme events. After having given the definitions, the tail dependence coefficients (TDC) are calculated for the copulae presented in the previous chapter. Then, an introduction into Extreme Value Theory (univariate and bivariate) is given, since it focuses on extreme values of random variables as we do in this chapter, as well. Then, the estimation of the TDC is discussed, for various assumptions. It turns out that it is important for the quality of the estimation of the TDC whether or not tail dependence is assumed. Therefore, a test for tail independence is presented, based on Extreme Value Theory.

3.1 Definition

The tail dependence coefficient is roughly speaking the probability that a random variable exceeds a certain threshold given that another random variable has already exceeded that threshold. More formally, the upper and lower TDC are defined as follows.

Definition 3.1.1. *Let $X = (X_1, X_2)^\top$ be a two dimensional random vector with marginal distribution functions F_1 and F_2. The coefficient of upper tail dependence of X is defined as:*

$$\lambda_U = \lim_{v \uparrow 1} P\{X_1 > F_1^{-1}(v) | X_2 > F_2^{-1}(v)\}.$$

Definition 3.1.2. *Analogously, the coefficient of lower tail dependence of X is defined as:*

$$\lambda_L = \lim_{v \downarrow 0} P\{X_1 \leq F_1^{-1}(v) | X_2 \leq F_2^{-1}(v)\}.$$

Definition 3.1.3. *We say that X is upper (lower) tail dependent if and only if $\lambda_U > 0$ ($\lambda_L > 0$). If $\lambda_U = 0$ ($\lambda_L = 0$) we say that X is upper (lower) tail-independent.*

3.2 Using copulae to estimate the TDC

The following proposition shows why the analysis of copulae is important when dealing with tail dependence.

Proposition 3.2.1. *The coefficient of upper tail dependence can be written in terms of copulae:*

$$\lambda_U = \lim_{v \uparrow 1} \frac{1 - 2v + C(v, v)}{1 - v},$$

where C is the copula of X. Analogously, we have

$$\lambda_L = \lim_{v \downarrow 0} \frac{C(v, v)}{v}.$$

Proof. Use the definition of the conditional expectation and Bayes' rule. □

In some cases, the TDC can be easily computed as shown in the next two examples.

Example 3.2.1. *For the Gumbel copula, we have*

$$C(v, v) = \exp[-\{(-\log v)^\theta + (-\log v)^\theta\}^{\frac{1}{\theta}}] = \exp(2^{\frac{1}{\theta}} \log v) = v^{2^{\frac{1}{\theta}}},$$

and therefore:

$$\lambda_U = \lim_{v \uparrow 1} \frac{1 - 2v + C(v, v)}{1 - v} = \lim_{v \uparrow 1} \frac{-2 + 2^{\frac{1}{\theta}} v^{2^{\frac{1}{\theta}}}}{-1} = 2 - 2^{\frac{1}{\theta}},$$

and

$$\lambda_L = \lim_{v \downarrow 0} \frac{C(v, v)}{v} = \lim_{v \downarrow 0} \frac{2^{\frac{1}{\theta}} v^{2^{\frac{1}{\theta}}}}{1} = 0.$$

Example 3.2.2. *For the Clayton copula, we have*

$$C(v, v) = (v^{-\theta} + v^{-\theta} - 1)^{-\theta^{-1}} = (2v^{-\theta} - 1)^{-\theta^{-1}},$$

and therefore:

$$\lambda_U = \lim_{v \uparrow 1} \frac{1 - 2v + C(v, v)}{1 - v} = \lim_{v \uparrow 1} \frac{-2 + 2v^{-\theta-1}(2v^{-\theta} - 1)^{-\theta^{-1}-1}}{-1} = 0,$$

and

$$\lambda_L = \lim_{v \downarrow 0} \frac{C(v, v)}{v} = \lim_{v \downarrow 0} (2 - v^\theta)^{-\theta^{-1}} = 2^{-\frac{1}{\theta}}.$$

This means that a Gumbel copula is able to model upper, whereas a Clayton copula can model lower tail dependence. Therefore, a combination of both can model asymmetric upper and lower tail dependence, as some might expect it for financial markets: losses occur more often jointly than gains do. The following example shows how to calculate the TDC for mixture copulae provided the TDCs are known for each of the copulae of the mixture.

Example 3.2.3. *The upper TDC for a mixture of two copulae is*

$$\begin{aligned}\lambda_U &= \lim_{v\uparrow 1}\frac{1-2v+C(v,v)}{1-v} = \lim_{v\uparrow 1}\frac{1-2v+\lambda C_1(v,v)+(1-\lambda)C_2(v,v)}{1-v}\\ &= \lambda\lim_{v\uparrow 1}\frac{1-2v+C_1(v,v)}{1-v} + (1-\lambda)\lim_{v\uparrow 1}\frac{1-2v+C_2(v,v)}{1-v}\\ &= \lambda\lambda_U^1 + (1-\lambda)\lambda_U^2,\end{aligned}$$

and analogously the lower TDC:

$$\begin{aligned}\lambda_L &= \lim_{v\downarrow 0}\frac{C(v,v)}{v} = \lim_{v\downarrow 0}\frac{\lambda C_1(v,v)+(1-\lambda)C_2(v,v)}{v}\\ &= \lambda\lambda_L^1 + (1-\lambda)\lambda_L^2,\end{aligned}$$

where λ_U^i and λ_L^i are the TDCs of copula i $(i = 1, 2)$.

We have the following relationship between the TDC for a copula C and the TDC of its survival copula $\tilde{C}$:

$$\lambda_U = \lim_{v\uparrow 1}\frac{1-2v+C(v,v)}{1-v} = \lim_{v\uparrow 1}\frac{\tilde{C}(1-v,1-v)}{1-v} = \lim_{t\downarrow 0}\frac{\tilde{C}(t,t)}{t},$$

since $\tilde{C}(1-v,1-v) = 1-2v+C(v,v)$. This means that the upper TDC of the copula is the lower TDC of its survival copula and vice versa.

For elliptically contoured copulae, the calculation is less straightforward than in the case of Archimedean copulae. The Gaussian copula is tail independent since its behavior in the tails is exponential and not proportional to a power law. For the t-copula, Embrechts et al. [2001] calculate the following TDC (since the t-distribution is symmetric, we have $\lambda_U = \lambda_L$):

Proposition 3.2.2. *The TDC for the t-copula is given by:*

$$\lambda_U = \lambda_L = 2 - 2t_{\nu+1}(\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho}),$$

where ν denotes the degrees of freedom, $t_\nu(z)$ the value of the t-distribution function with ν degrees of freedom at point z and ρ is the correlation parameter (which is for $\nu > 2$ the coefficient of linear correlation).

Proof. See Embrechts et al. [2001]. □

Table 3.1 summarizes the TDC for the copulae discussed in the previous section.

Copula	λ_U	λ_L
Clayton	0	$2^{-\frac{1}{\theta}}$
Plackett	0	0
Gumbel	$2-2^{\frac{1}{\theta}}$	0
t-copula	$2-2t_{\nu+1}(\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho})$	
Gaussian	0	
Mixt. Clayton/Gumbel	$(1-\lambda)(2-2^{\frac{1}{\theta_2}})$	$\lambda 2^{-\frac{1}{\theta_1}}$
Mixt. Clayton/S-Clayton	$(1-\lambda)2^{-\frac{1}{\theta_2}}$	$\lambda 2^{-\frac{1}{\theta_1}}$
Mixt. S-Clayton/S-Gumbel	$\lambda 2^{-\frac{1}{\theta_1}}$	$(1-\lambda)(2-2^{\frac{1}{\theta_2}})$
Mixt. Gumbel/S-Gumbel	$\lambda(2-2^{\frac{1}{\theta_1}})$	$(1-\lambda)(2-2^{\frac{1}{\theta_2}})$

Table 3.1: TDCs for different copulae

From table 3.1, it is obvious that when we assume a certain copula, estimation of the TDC is straightforward: The parameters of the copula have to be replaced by their estimators to obtain the estimator of the TDC. But assuming a certain copula can be restrictive and estimation under misspecification misleading. Therefore, Extreme Value Theory is introduced in the next section, which allows for estimation of the TDC under less restrictive assumptions.

For the estimation of the TDC in elliptically contoured copulae models, see e.g. Klüppelberg et al. [2006].

3.3 Extreme Value Theory

Since analyzing tail dependence is all about extreme values, this section presents results from univariate and bivariate Extreme Value Theory (EVT), which are then used in the context of copulae and tail dependence. It shows that, under some assumptions, estimators for the TDC can be derived using EVT, which are less restrictive than the estimators assuming a parametric copula model.

3.3.1 Univariate Extreme Value Theory

In order to derive results for bivariate extreme values, a short introduction in univariate Extreme Value Theory is given.

Let $X_n^* = \max(X_1, \ldots, X_n)$ and $X_n^- = \min(X_1, \ldots, X_n)$ be the maximum and minimum respectively of a random sample $(X_1, \ldots, X_n)$ of independent copies of X. Now define for some sequences $a_n \in \mathbb{R}_+^{\mathbb{N}}$, $b_n \in \mathbb{R}^{\mathbb{N}}$ the standardized maximum by:

$$\tilde{X}_n^* = \frac{X_n^* - b_n}{a_n}.$$

Via $\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n)$, all the following results for maxima can be applied for minima, too.

Now, in definition 3.3.1, the so called family of Generalized Extreme Value distributions is given. This family will show in theorem 3.3.1 to be the only possible limiting distribution of standardized maxima.

Definition 3.3.1. *The family of Generalized Extreme Value (GEV) distributions is given by:*

$$G(x) = \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi} \right\},$$

where $(x)_+ = \max(x, 0)$, $\sigma \in \mathbb{R}_+$ *and* $\xi, \mu \in \mathbb{R}$.

Remark 3.3.1. *In this definition,* μ, σ *and* ξ *are location, scale and shape parameter respectively. For simplicity of notation, in the following we often standardize to* $\mu = 0$ *and* $\sigma = 1$.

Theorem 3.3.1. *(Unified extremal types theorem, UETT): If there exists sequences* $a_n \in \mathbb{R}_+^{\mathbb{N}}$, $b_n \in \mathbb{R}^{\mathbb{N}}$ *such that*

$$P\left(\frac{X_n^* - b_n}{a_n} \leq x\right) = P(\tilde{X}_n^* \leq x) \stackrel{n\to\infty}{\longrightarrow} G(x),$$

for a non degenerated distribution function G*, then* G *belongs to the GEV family:*

$$G(x) = \exp\left\{ -(1 + x\xi)_+^{-1/\xi} \right\}.$$

Proof. For the proof, see e.g. Ferreira and de Haan [2006], p.7-8. For an idea of the proof, see e.g. Coles [2001], p.49-51. □

Remark 3.3.2. *For* $\xi \to 0$ *we obtain the Gumbel distribution:* $H_0(x) = \exp\{-e^{-x}\}$*, for* $\xi > 0$ *the Frechet distribution* $H_1(1+\xi x, 1/\xi)$*, where* $H_1(x,\theta) = \exp\{-x^{-\theta}\}$ *and for* $\xi < 0$ *we obtain the Weibull distribution* $H_{-1}(-1-\xi x, -1/\xi)$ *where,* $H_{-1}(x,\theta) = \exp\{-(-x)^{\theta}\}$.

The following definition introduces the term of domain of attraction, which is often used in the literature.

Definition 3.3.2. *Domain of attraction: if*

$$P(\tilde{X}_n^* \leq x) \stackrel{n\to\infty}{\longrightarrow} G(x),$$

where G *belongs to the GEV, we say that* F*, the distribution function of* X*, is in the domain of attraction of* G.

In order to derive necessary and sufficient conditions for distributions to be in the domain of attraction of a GEV, we define the reciprocal hazard function:

Definition 3.3.3. *For the distribution function F and its derivative f, the reciprocal hazard function is given by.*

$$h(x) = \frac{1 - F(x)}{f(x)}.$$

Using this function h, a sufficient condition for a GEV can be stated:

Theorem 3.3.2. *(von Mises' condition) If*

$$h'(x) \overset{x \to x^*}{\longrightarrow} \xi,$$

F is in the domain of attraction of a GEV with parameter ξ. a_n and b_n are chosen as follows: $1 - F(b_n) = 1/n$ and $a_n = h(b_n)$.

Proof. See e.g. Ferreira and de Haan [2006], p.15-16. □

Example 3.3.1. *For the exponential distribution with parameter 1 and cdf $F(x) = 1 - e^{-x}$, we get $h(x) = \frac{e^{-x}}{e^{-x}} = 1$ and hence $h'(x) = 0, \forall x$. Therefore, by choosing $b_n = \log(n)$ and $a_n = 1$, $X_n^* - \log n$ converges to a Gumbel distribution.*

Remark 3.3.3. *Ferreira and de Haan [2006] give necessary and sufficient conditions for a distribution function being in the domain of attraction of a GEV. Since most commonly known distributions are in the domain of attraction of a GEV, these conditions are skipped and we will focus on aspects that are more relevant in our context.*

Now, focus on a different aspect of EVT: the convergence of a the distribution over a (sufficiently high chosen) threshold, which involves the notion of Generalized Pareto distribution:

Theorem 3.3.3. *Suppose that X fulfills the conditions of the UETT, i.e.:*

$$P(\tilde{X}_n^* \leq x) \overset{n \to \infty}{\longrightarrow} G(x) = \exp\left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\},$$

then the distribution of $Y_i = X_i - k | X_i > k$ converges for $k \to \infty$ to:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}} \right)_+^{-1/\xi}$$

for $y > 0$, where $\tilde{\sigma} = \sigma + \xi(k - \mu)$. The family of distributions, which have H as cdf is called Generalized Pareto family.

Proof. See Coles [2001], p. 76-77. □

3.3.2 Bivariate Extreme Value Theory

The discussion of bivariate extreme value and dependence concepts based on it, emerged rather recently, see e.g. citeptawn88. In the univariate case, working with maxima is intuitive. In the bivariate or higher dimensional case, it is less trivial to define the maximum (of a vector). One approach is to define component-wise maxima.

For a good overview on multidimensional extreme value models, see Kotz and Nadarajah [2000].

Definition 3.3.4. *Define for a random sample* $(X_{11}, \ldots, X_{1n}), (X_{21}, \ldots, X_{2n})$ *of independent copies of* (X_1, X_2) *the component-wise maxima and minima:*

$$X^*_{jn} = \max((X_{j1}, \ldots, X_{jn})),\ j = 1, 2$$

and

$$X^-_{jn} = \min((X_{j1}, \ldots, X_{jn})),\ j = 1, 2$$

and note the vectors of component-wise maxima and minima by $X^*_n = (X^*_{1n}, X^*_{2n})$ *and* $X^-_n = (X^-_{1n}, X^-_{2n})$ *respectively.*

For the component-wise maxima, the results of univariate Extreme Value Theory apply. Put differently, the marginal distributions being in the domain of attraction of a GEV is a necessary condition for that the component-wise maxima can be in the domain of attraction of a (later defined) bivariate GEV. To simplify the presentation, Coles [2001] assumes that X^*_{1n} and X^*_{2n} have standard Frechet distribution, i.e. $\forall x > 0$:

$$F_{X^*_{1n}}(x) = F_{X^*_{2n}}(x) = e^{-1/x}.$$

The following theorem 3.3.4 generalizes the UETT from theorem 3.3.1 to the bivariate case.

Theorem 3.3.4. *Assuming that* X^*_{1n} *and* X^*_{2n} *have standard Frechet distribution and that there exist sequences* $a_n \in \mathbb{R}^{\mathbb{N}}_+$, $b_n \in \mathbb{R}^{\mathbb{N}}$ *such that*

$$P(\tilde{X}^*_{1n} < x_1, \tilde{X}^*_{2n} < x_2) \overset{n \to \infty}{\longrightarrow} G(x_1, x_2),$$

for a non degenerated distribution function G*, then* G *belongs to the bivariate extreme value distributions:*

$$G(x_1, x_2) = \exp(-V(x_1, x_2)),$$

where

$$V(x_1, x_2) = 2 \int_0^1 \max\left(\frac{w}{x_1}, \frac{1-w}{x_2}\right) dH(w)$$

and H *is a distribution function on* $[0, 1]$ *satisfying:*

$$\int_0^1 w dH(w) = 1/2.$$

Proof. See e.g. Ferreira and de Haan [2006], p. 211. □

Remark 3.3.4. *By plugging in* $\bar{x}_i = \left[1 + \xi_i\left(\frac{x_i - \mu_i}{\sigma_i}\right)\right]_+^{1/\xi_i}$ *for* $i = 1, 2$ *in the above theorem instead of* x_1 *and* x_2*, one obtains the general (not assuming Frechet margins) version of the bivariate extreme value distribution, see Coles [2001], p. 145.*

Now we would like to find an equivalent of theorem 3.3.3 for the bivariate case.

Theorem 3.3.5. *Assuming that* X_1 *and* X_2 *can be approximated by a Generalized Pareto Distribution, we define:*

$$\tilde{x_i} = -\left[\log\left\{1 - \zeta_i\left(1 + \frac{\xi_i(X_i - k_i)}{\sigma_i}\right)^{-1/\xi_i}\right\}\right]^{-1}, \quad i = 1, 2$$

Then the joint distribution function of $x_1 > k_1$ *and* $x_2 > k_2$ *is approximately:*

$$G(x_1, x_2) = \exp\{-V(\tilde{x_1}, \tilde{x_2})\}.$$

Proof. See Coles [2001], p. 154-155. □

3.4 Applying EVT to copulae and tail dependence

Now, we want to write the above results of Extreme Value Theory in terms of copulae and apply the results to the estimation problem of tail dependence. Some further definitions are needed.

Proposition 3.4.1. *The following relation exists between the copula* C_l^* *of the component-wise maxima* $(X_{1l}^*, X_{2l}^*), l \in \mathbb{N}$ *and the copula* C *of* (X_1, X_2)*:*

$$C_l^*(u_1, u_2) = C^l(u_1^{1/l}, u_2^{1/l}).$$

Proof. See Nelsen [2006], p. 95. □

Remark 3.4.1. *This proposition can be extended to hold for all* $t > 0$ *instead of* $l \in \mathbb{N}$*, see Joe [1997], p. 173.*

Remark 3.4.2. *If* $C(u_1, u_2) = C^l(u_1^{1/l}, u_2^{1/l})$ *holds* $\forall u, v \in [0, 1], \forall l > 0$*,* C *is said to be max-stable.*

Definition 3.4.1. *A copula* C^* *is an extreme value copula if there exists a copula* C *such that:*

$$C^*(u_1, u_2) = \lim_{n \to \infty} C^n(u_1^{1/n}, u_2^{1/n}).$$

Theorem 3.4.1. *A copula is max-stable if and only if it is an extreme value copula.*

Proof. See Nelsen [2006], p.97. □

Let us now come back to the main result of EVT, the Generalized Extreme Value distribution. Pickands [1981] introduced a representation, which is now widely used. He assumes standard exponential margins ($\lambda = 1$ and hence $\bar{F}_1(0, x) = \bar{F}_2(x, 0) = e^{-x}$) and shows that then the joint survival function is given by $\bar{F}(x_1, x_2) = \exp[-(x_1 + x_2)D(\frac{x_1}{x_1+x_2})]$ for some function D. The following theorem summarizes this approach (see Joe [1997], p.175).

Theorem 3.4.2. *Let D a continuous, non negative function on $[0,1]$ with $D(0) = D(1) = 1$, twice left and right differentiable except for at most a countable number of points. Then the following two assertions are equivalent:*

1. $G(x_1, x_2) = \exp[-(x_1 + x_2)D(\frac{x_1}{x_1+x_2})]$ *is a bivariate exponential survival function.*
2. *D is convex with* $\max(z, 1-z) \leq D(z) \leq 1, \forall z \in [0,1]$.

We call D Pickands dependence function.

Proof. See Joe [1997], p.175-176 □

Remark 3.4.3. *If $X \sim \exp(1)$ then $Y = 1/X \sim$ Frechet with cdf $F(x) = e^{-1/x}$. Then the above theorem is the same, except for $G(x_1, x_2) = \exp[(x_1 + x_2)D(\frac{x_1}{x_1+x_2})]$.*

In terms of copulae, this result becomes: if C is an extreme value copula (using Frechet margins), it is of the form:

$$C_{EV}(u_1, u_2) = \exp\left[\log(x_1x_2)D\left(\frac{\log(x_1)}{\log(x_1x_2)}\right)\right].$$

Using this fact, we can calculate $C_{EV}(v, v)$:

$$C_{EV}(v, v) = F_{EV}(\log(v), \log(v)) = \exp(2\log(v)D(1/2)).$$

Now assume that the conditions for the Generalized Extreme Value theorem hold. Since,

$$\begin{aligned}
\lambda_U &= \lim_{v\uparrow 1} \frac{1 - 2v + C_{EV}(v,v)}{1-v} = 2 - \lim_{v\uparrow 1} \frac{1 - C_{EV}(v,v)}{1-v} \\
&= 2 - \lim_{v\uparrow 1} \frac{dC_{EV}(v,v)}{dv} = 2 - \lim_{v\uparrow 1} 2D(1/2)\exp(2\log(v)D(1/2)) \\
&= 2 - 2D(1/2),
\end{aligned}$$

we can estimate λ_U by estimating D. Furthermore, we see that $D(z) = 1, z \in [0,1]$ characterizes tail independence.

3.5 Different concepts and notations for the dependence function and the TDC

Many different notations concerning EVT and tail dependence exist in the literature. This section presents the concepts of different authors and illuminates the link between them (see also Heffernan [2000]). Overall, Pickands dependence function is the most utilized concept. In theorem 3.3.5, we introduced a function V, which is linked to Pickands function via the following relation:

$$D(t) = V\left(\frac{1}{1-t}, \frac{1}{t}\right).$$

Ferreira and de Haan [2006] work with the exponent measure ν and a function L, defined by:

$$L(1-t, t) = D(t) = V\left(\frac{1}{1-t}, \frac{1}{t}\right).$$

Therefore, estimation of L is the same as estimation of D or V. Another representation stems from Ledford and Tawn [1996, 1997, 1998]: With unit Frechet margins and under broad conditions, the following relation holds:

$$\mathrm{P}(X_1 > k, X_2 > k) \sim \mathcal{L}(t)(\mathrm{P}(X_1 > k))^{1/\eta},$$

where $\mathcal{L}$ is a slowly varying function, i.e. for fixed x, $\lim\limits_{t\to\infty} \dfrac{\mathcal{L}(tx)}{\mathcal{L}(t)} = 1$. They call η the coefficient of tail dependence. But since η is not equal to λ_U, we will refer to η as the residual dependence index, as do Ferreira and de Haan [2006].

Coles et al. [1999] and Coles [2001] uses a different notation, yet. He defines $\chi(v) = 2 - \frac{\log C(v,v)}{\log v}$, and its limit $\chi = \lim\limits_{v\uparrow 1} \chi(v)$, which is equal to the definition of the upper TDC above ($\chi = \lambda_U$). Furthermore, he introduces $\bar{\chi}(v) = \frac{2\log(1-v)}{\log \tilde{C}(v,v)} - 1$ and $\lim\limits_{v\uparrow 1} \bar{\chi}(v) = \bar{\chi}$. We get the following characterization of tail dependence and tail independence:

$$\begin{cases} \chi > 0, \bar{\chi} = 1 & \text{asymptotic dependence,} \\ \chi = 0, \bar{\chi} < 1 & \text{asymptotic independence.} \end{cases}$$

The relation between the TDC and the notation of Ledford and Tawn is the following one: $\bar{\chi} = 2\eta - 1$ and

$$\chi = \begin{cases} c & \text{if } \bar{\chi} = 1 \text{ and } L(t) \to c > 0 \text{ for } t \to \infty, \\ 0 & \text{if } \bar{\chi} = 1 \text{ and } L(t) \to 0 \text{ for } t \to \infty, \\ 0 & \text{if } \bar{\chi} < 1. \end{cases}$$

3.6 Estimation of the TDC

Now, there are two possibilities to use Extreme Value Theory for the estimation of the TDC (particularly for financial data). The first one is to develop estimators based on the assumptions of the Generalized Pareto Distribution (GPD). Therefore, one assumes convergence (over some threshold) to a bivariate Generalized Pareto Distribution. Then, the dependence function is estimated. This model is called Peaks over threshold or threshold exceedances model. The other possibility is to assume that the assumptions of the GEV are fulfilled. This will (in a financial application) rarely be the case. Therefore, we use block-maxima. Both methods come to the same estimation problem: the dependence function is to be estimated. The difference is the treatment of the data: In the first case, we choose the realizations that lie above a threshold, in the second case block-maxima.

3.6.1 Peaks over threshold models

Ferreira and de Haan [2006] p. 236 present an estimator for Pickands dependence function (or in their notation L). Recall the link between the TDC and L: $\lambda_U = 2 - 2D(1/2) = 2 - 2L(1/2, 1/2)$. Assume that the sample is ordered and let $X_{k,n}$ denote the kth upper order statistic. Then Ferreira and de Haan's estimator for L is:

$$\hat{L}(x_1, x_2) = \frac{1}{k} \sum_{i=1}^{n} \mathrm{I}\{X_{1i} \geq X_{1,n-[kx]+1,n} \text{ or } X_{2i} \geq X_{2,n-[kx]+1,n}\},$$

where [x] is the floor function. For $t = 1/2$, we get:

$$\hat{L}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{k} \sum_{j=1}^{n} \mathrm{I}(R_{n1}^{(j)} \geq n - k \text{ or } R_{n2}^{(j)} \geq n - k),$$

where $R_{n1}^{(j)}$ and $R_{n2}^{(j)}$ are the ranks of $X_1^{(j)}$ and $X_2^{(j)}$, $\forall j = 1, \ldots, n$. This estimator is not necessarily convex; therefore the authors introduce another estimator, which uses an estimator of the spectral density. The following estimator is convex. But, as can be easily seen, for $L(1/2, 1/2)$, both coincide.

$$\hat{L}_{\hat{\Phi}}(x, y) = x\hat{\Phi}(\frac{\pi}{2}) + \max(x, y) \int_{\pi/4}^{\arctan(y/x)} \hat{\Phi}(\theta) \left\{ \max\left(\frac{1}{\sin^2\theta}, \frac{1}{\cos^2\theta}\right) \right\} d\theta,$$

where

$$\hat{\Phi}(\theta) = \frac{1}{k} \sum_{j=1}^{m} \mathrm{I}\{\max(R_{m1}^{(j)}, R_{m2}^{(j)}) \geq m+1-k \text{ and } n+1-R_{m2}^{(j)} \leq (n+1-R_{m1}^{(j)}) \tan\theta\}.$$

Obviously,

$$\hat{L}_{\hat{\Phi}}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2}\hat{\Phi}\left(\frac{\pi}{2}\right) = \hat{L}\left(\frac{1}{2}, \frac{1}{2}\right).$$

Therefore, in the following, only $\hat{L}$ is used whose asymptotic properties are summarized in the following theorem:

Theorem 3.6.1. *$\hat{L}$ is a consistent and asymptotically normal estimator of L.*

Proof. see Ferreira and de Haan [2006], p. 237-244. □

Now an estimator of λ_U is given by:

$$\hat{\lambda}_{U,n}^{(1)} = 2 - 2\hat{L}(1/2, 1/2).$$

Defining the empirical copula and plugging in this as the empirical counterpart of the copula in the definition of the TDC can also motivate this estimator.

Definition 3.6.1. *The empirical copula is defined as:*

$$\hat{C}_n(u,v) = \frac{1}{n}\sum_{j=1}^{n} I\left(\frac{R_{1j}}{n} \leq u, \frac{R_{2j}}{n} \leq v\right)$$

where F_n, F_{1n} and F_{2n} denote the empirical cdf corresponding to F, F_1 and F_2 respectively.

Therefore, the above estimator is equal to:

$$\begin{aligned}
\hat{\lambda}_{U,n}^{(1)} &= \frac{n}{k}\hat{C}_n\left(\left(1-\frac{k}{n},1\right]\times\left(1-\frac{k}{n},1\right]\right) \\
&= \frac{1}{k}\sum_{j=1}^{n} \mathrm{I}(R_{n1}^{(j)} > n-k, R_{n2}^{(j)} > n-k),
\end{aligned}$$

where $k = k(n) \overset{n\to\infty}{\longrightarrow} \infty$ and $\frac{k}{n} \overset{n\to\infty}{\longrightarrow} 0$. See Schmidt [2005] for details.

Using the notation of Coles, two estimators can be used as well. Recall that

$$\frac{1-2v+C(v,v)}{1-v} = 2 - \frac{1-C(v,v)}{1-v} \overset{v\uparrow 1}{\sim} 2 - \frac{\log C(v,v)}{\log v} = \chi(v),$$

we define the following two estimators:

$$\hat{\lambda}_{U,n}^{(2)} = 2 - \frac{\log \hat{C}_n((n-k)/k, (n-k)/k)}{\log((n-k)/k)},$$

and

$$\hat{\lambda}_{U,n}^{(3)} = 2 - \frac{1-\hat{C}_n((n-k)/k, (n-k)/k)}{1-(n-k)/k},$$

where k is a threshold (to be chosen).

Now, we have three different estimators based on peaks over threshold models. Let us come to the other possibility of EVT to estimate the TDC: block-maxima models.

3.6.2 Block-maxima models

Using block-maxima models means using the Generalized Extreme Value distribution. Since most data is not the maximum of something, we use block-maxima, i.e. the maximum of say, 5 or ten observations. Details will be given in chapter 4. Capéraà et al. [1997] present an estimator $A_n(t)$ for Pickands dependence function $A(t)$. They use the so called Pickands coordinates (see Falk and Reiss [2003], Falk and Reiss [2005a] and Falk and Reiss [2005b] for details), i.e. $Z_i = \log(U_i)/\log(U_iV_i)$ with cdf $H(z) = \mathrm{P}(Z_i \leq z) = z + (1-z)D(z)$ where $D(z) = A'(z)/A(z)$, with A' the right derivative of A for all $z \in [0,1)$. The estimator is given by:

$$A_n(t) = \begin{cases} (1-t)Q_n^{1-p(t)} & \text{if } 0 \leq t \leq Z_{\{1\}}, \\ t^{i/n}(1-t)^{1-i/n}Q_n^{1-p(t)}Q_i^{-1} & \text{if } Z_{\{i\}} \leq t \leq Z_{\{i+1\}},\ i = 1,\ldots,n-1, \\ tQ_n^{-p(t)} & \text{if } Z_{\{n\}} \leq t \leq 1, \end{cases}$$

where p is a weighting function and:

$$Q_i = \left\{ \prod_{k=1}^{i} Z_{\{k\}}/(1-Z_{\{k\}}) \right\}^{1/n}, \quad i = 1,\ldots,n-1,$$

and $Z_{\{k\}}$ denoting the ordered sample of $Z_i,\ i = 1,\ldots,n$. The following theorem summarizes the properties of the estimator.

Theorem 3.6.2. *Let p a bounded function on $[0,1]$. A_n is an asymptotically unbiased estimator of A, which is uniformly strongly consistent.*

Proof. Capéraà et al. [1997]. □

Capéraà et al. [1997] choose $p(t) = 1-t$. Therefore, for estimating $A(1/2)$ we obtain the following estimator:

$$A_n(1/2) = \frac{1}{2}Q_n^{1/2}Q_i^{-1},$$

where:

$$Q_i = \left\{ \prod_{k=1}^{i} Z_{\{k\}}/(1-Z_{\{k\}}) \right\}^{1/n}, \quad i = 1,\ldots,n-1,$$

and the estimator for the upper TDC becomes:

$$\hat{\lambda}_{U,n}^{(4)} = 2 - 2A_n(1/2),$$

which is equal to the following estimator in Frahm et al. [2005]:

$$\hat{\lambda}_U^{CFG} = 2 - 2\exp\left[\frac{1}{n}\sum_{i=1}^{n}\log\left\{\sqrt{\log\frac{1}{U_i}\log\frac{1}{V_i}} \Big/ \log\frac{1}{\max(U_i,V_i)^2}\right\}\right].$$

In the notation of Frahm et al. [2005], the presented estimators are: $\hat{\lambda}_U^{(2)} = \hat{\lambda}_U^{LOG}$ and $\hat{\lambda}_U^{(3)} = \hat{\lambda}_U^{SEC}$. Frahm et al. [2005] give furthermore estimators for the TDC under different assumptions: Using a specific distribution (e.g. t-distribution), within a class of distributions (e.g. elliptically contoured distributions), using a specific copula (e.g. Gumbel copula), within a class of copulae (e.g. Archimedean copulae) or a nonparametric estimation (without any parametric assumption). By means of a simulation study, the authors compare the performance of the different estimators for different cases: whether the assumption is true or wrong and whether there is tail dependence or not. It turns out that what Frahm et al. [2005] call the "nonparametric" estimators ($\hat{\lambda}^{CFG}, \hat{\lambda}^{SEC}, \hat{\lambda}^{LOG}$) perform well if there is tail dependence but bad if there is not. The estimators assuming a specific copula, copula class or distribution class perform well if the assumption is correct but bad if the assumption is wrong. That means we choose the estimators according to the prior information. If reasonable assumptions about the copula or the distribution could be made, one could use a specific copula or distribution class. In practical applications, one will never know which copula model is the correct one. The estimation can only be under misspecification. Since it is difficult to select a copula model, based on a statistical criterion, we will try to work with weaker assumptions and use the estimators $\hat{\lambda}^{(1)}, \hat{\lambda}^{(2)}, \hat{\lambda}^{(3)}$ and $\hat{\lambda}^{(4)}$. But, as mentioned above, these estimators for the TDC perform well if tail dependence exists in the data, but bad if there is not. This brings us to the important issue of testing for tail dependence.

3.7 Testing for tail independence

3.7.1 A test based on the residual dependence index

Draisma et al. [2004] present three different estimators for the residual dependence index η: a Maximum likelihood estimator in a Generalized Pareto model, a Hill estimator and the estimator presented in Peng [1999]. Here, the Hill estimator is presented since it can be easily implemented and as Draisma et al. [2004] argue, the ML's advantage of location invariance over the Hill estimator is not relevant here (after standardization of the margins). Furthermore the Hill estimator has lower variance. Peng's estimator turns out to be outperformed by the other two estimators in the simulation study by Draisma et al. [2004]. The Hill estimator is defined as:

$$\hat{\eta} = \frac{1}{m} \sum_{i=1}^{m} \log \frac{T_{n,n-1+i}^{(n)}}{T_{n,n-m}^{(n)}},$$

where $T_{n,k}^{(n)}$ is the k-th order statistic of

$$T_i^{(n)} = \min\left(\frac{n+1}{n+1-R_i^{X_1}}, \frac{n+1}{n+1-R_i^{X_2}} \right).$$

Theorem 3.7.1. *Under some conditions (for details see Draisma et al. [2004]),* $\sqrt{m}(\hat{\eta} - \eta)$ *is asymptotically normal with mean* 0 *and variance:*

$$\sigma^2 = \eta^2(1 - l)(1 - 2lc_{X_1}(1,1)c_{X_2}(1,1)).$$

Proof. see Draisma et al. [2004]. ☐

The estimator for σ^2 is given in the following theorem. Before, we define estimators for the unknown parts of σ^2. Let $\hat{l} = \frac{m}{n}T^{(n)}_{n,n-m}$ the estimator of

$$l = \lim_{v \downarrow 0} = t^{-1}\mathrm{P}(1 - F_1(X_1) < t, 1 - F_2(X_2) < t),$$

and let $\hat{k} = m/\hat{l}$. Furthermore, define

$$\hat{c}_{X_1}(1,1) = \frac{\hat{k}^{5/4}}{n}(T^{(n,\hat{k}^{1/4})}_{n,n-m} - T^{(n)}_{n,n-m}),$$

where $T^{(n,u)}_{n,i}$ is the i-th order statistics of:

$$T^{(n,u)}_i = \min\left(\frac{n+1}{n+1-R^{X_1}_i}(1+u), \frac{n+1}{n+1-R^{X_2}_i}\right).$$

Define $\hat{c}_{X_2}(1,1)$ analogously to $\hat{c}_{X_1}(1,1)$.

Theorem 3.7.2. *The variance estimator:*

$$\hat{\sigma}^2 = \hat{\eta}^2(1 - \hat{l})(1 - 2\hat{l}\hat{c}_{X_1}(1,1)\hat{c}_{X_2}(1,1))$$

is consistent, $\forall \eta \in (0,1]$.

Proof. see Draisma et al. [2004]. ☐

Using this, a one-sided test of tail independence (null hypothesis: η<1) can be easily constructed.

3.7.2 A different approach for testing for tail independence

Another approach for testing for tail independence is given in Falk and Michel [2006]. They prove the following theorem.

Theorem 3.7.3. *With* $c \uparrow 0$*, we have uniformly for* $t \in [0,1]$*:*

$$P(X_1+X_2 > ct | X_1+X_2 > c) = \begin{cases} t^2(1+O(c)) \text{ if there is no tail dependence,} \\ t(1+O(c)) \text{ else.} \end{cases}$$

Proof. See Falk and Michel [2006]. ☐

Using this theorem, Falk and Michel propose four different tests for tail independence, which can be grouped into 2 different classes: a log-likelihood ratio (LR) test and three goodness of fit tests (Fisher's κ, Kolmogorov-Smirnov and χ^2). In the latter class, the Komolgorov-Smirnov-test (KS) turns out to be the best in the simulation study by Falk and Michel [2006]. Therefore, in the following, only LR and KS tests are described.

Likelihood Ratio test

Assume we have a random sample $(X_{11}, \dots, X_{1n}), (X_{21}, \dots, X_{2n})$ of independent copies of (X_1, X_2). The marginal distribution is assumed to be reverse exponential (i.e. $F(0, x) = F(x, 0) = e^x$). Now, fix a threshold $c < 0$ and consider $E = \{C_i = X_{1i} + X_{2i} : C_i > c\}$. Let $K(n) = \#E$ and define $V_i = C_i/c, \forall i = 1, ..., K(n)$.

The LR test considers the distribution function of V_i and tests whether it is more likely from $F_{(0)}(t) = t^2$ or $F_{(1)}(t) = t$. The test statistic for testing $F_{(0)}$ (tail independence) against $F_{(1)}$ is (for fixed n):

$$T_{LR}(V_1, \dots, V_{K(n)}) = -\sum_{i=1}^{K(n)} \log(V_i) - K(n)\log(2).$$

$F_{(0)}$ is rejected when T_{LR} gets large. The p-value is given by:

$$\begin{aligned} p_{LR} &= \exp(-2(T_{LR}(V_1, \dots, V_{K(n)})) + K(n)\log(2)) \\ &\quad \times \sum_{j=0}^{K(n)-1} \frac{(-2(T_{LR}(V_1, \dots, V_{K(n)})) + K(n)\log(2))^j}{j!} \\ &\approx \Phi\left(\frac{2\sum_{j=1}^{K(n)} \log(V_i) + K(n)}{K(n)^{1/2}}\right). \end{aligned}$$

Kolmogorov Smirnov goodness of fit test

A different possibility of using theorem 3.7.3 is to carry out a goodness-of-fit test, in this case using the Kolmogorov Smirnov test. Therefore, define, conditional on $K(n) = m$:

$$U_i = F_c(C_i/c) = \frac{1 - (1 - C_i)\exp(C_i)}{1 - (1 - c)\exp(c)}, \forall i \in \{1, \dots, m\}.$$

Denote $\hat{F}_m(t) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{I}_{[0,t]}(U_i)$ the ecdf of $U_i, i = 1, \dots, m$. The Kolmogorov test statistic is then:

$$T_{KS} = \frac{1}{m}\sup_{t \in [0,1]} |\hat{F}_m(t) - t|.$$

The approximate p-value is $p_{KS} = 1 - K(T_{KS})$, where K is the cdf of the Kolmogorov distribution. According to a rule of thumb given by the authors:

for $m > 30$, tail independence is rejected if $T_{KS} > c_{0.05} = 1.36$.

Since choosing c is difficult in practice and due to some technical problems, this test is omitted for the following analysis. It could be interesting in the future to carry out an extensive simulation study to compare the results of both tests. But due to the limits in time of a diploma thesis, only the test using the residual dependence index is implemented in the following chapters. In the next chapter, the small sample behavior of this test is assessed. In chapter 5, the test is applied to empirical financial data.

Chapter 4

Simulation

In the previous sections, we have seen the asymptotic results for the proposed estimators. Since stability of the parameters cannot be assumed over long time spans, one has to treat the tradeoff of choosing a window for estimation that can be reasonably assumed to have no variation of the parameters and to choose a long window length in order to reduce bias of the estimators. In this section, the small sample behavior of the different estimators and tests is assessed: for different thresholds and different block-maxima respectively, and for different distributions, simulations are carried out.

4.1 Estimation of the residual dependence index under various distributions

First of all, let us look at the estimation of the residual dependence index η. The estimation of this parameter enables us to test for tail independence in a Generalized Pareto Model. Figure 4.1 presents the estimation results for $\hat{\eta}$ for some elliptically contoured distributions with different parameters, figure 4.2 gives some simulation results for the Gumbel and Clayton copula. The full line always represents the case of tail dependence ($\hat{\eta} = 1$), the dotted line corresponds to the mean of the estimation over 1,000 simulations, the dash-dotted lines are the confidence intervals bounds (at 10%). Therefore, in a one-sided test, the null hypothesis of tail-independence can be rejected if the upper dash-dotted line is above the full line, i.e. 1. Recall, that except for the Gaussian distribution, all distributions have tail dependence.

The graphs illustrate the variance-bias trade-off in the estimation procedure. The higher the threshold, the lower the variance but the higher the bias. Therefore, the threshold has to be chosen approximately such that for the Gaussian distribution we can accept the Null hypothesis of tail independence, whereas for all others, we are able to reject it.

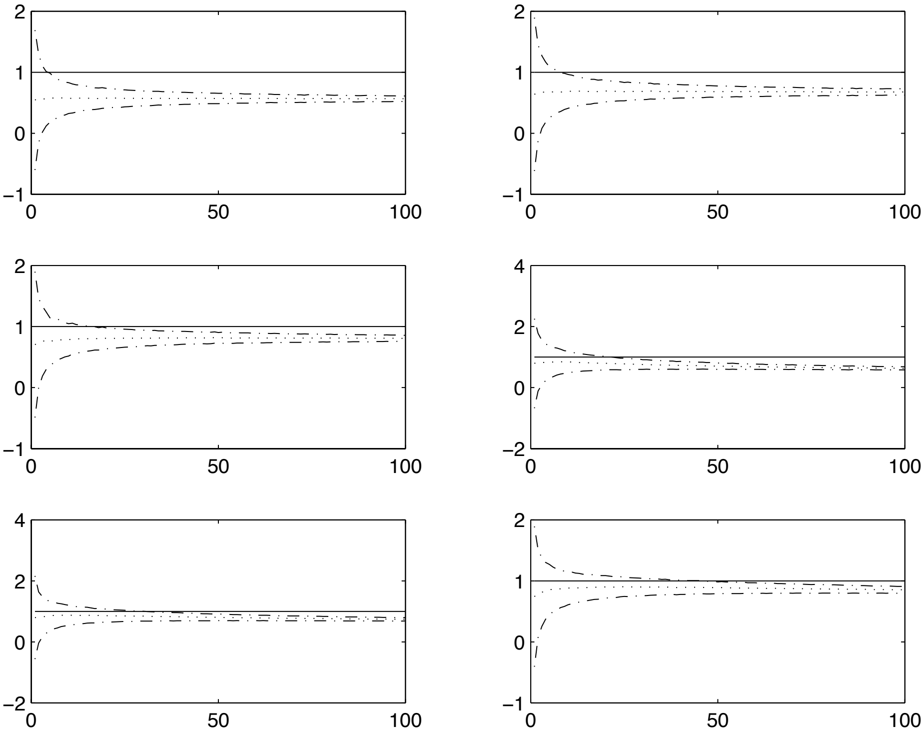

Figure 4.1: Estimation results of $\hat{\eta}$ (dotted) and confidence intervals (10%, dash-dotted) on the y-axis and m on the x-axis, for Gaussian ($\rho = 0.2$, $\rho = 0.5$ and $\rho = 0.8$), and t ($\rho = 0.2$, $\rho = 0.5$ and $\rho = 0.8$ ($\nu = 2$)); from left to right and top to bottom.

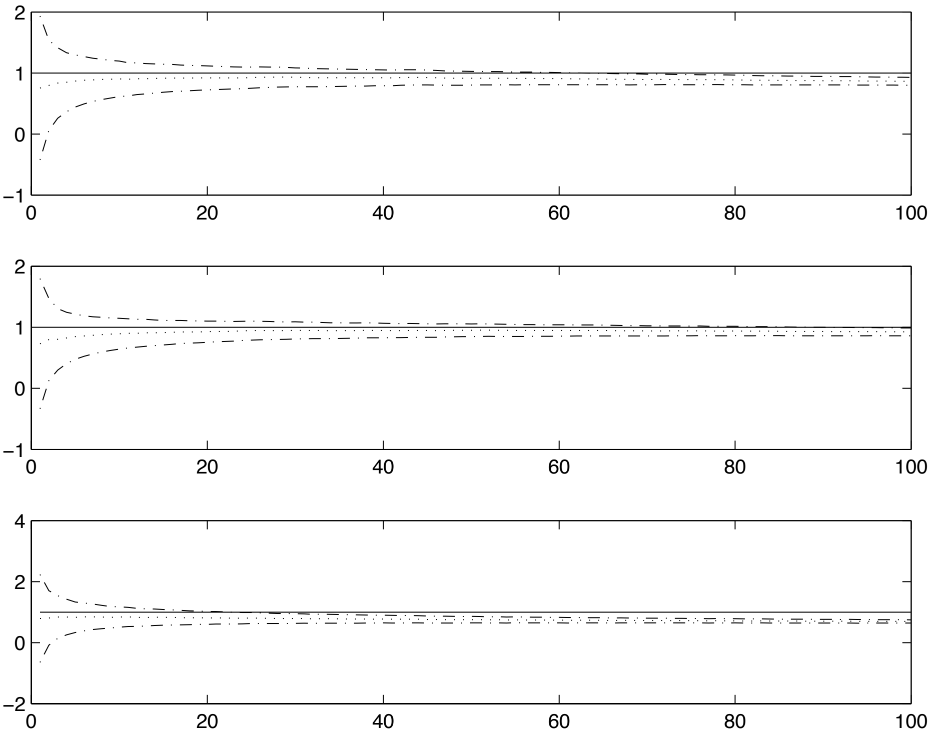

Figure 4.2: Estimation results of $\hat{\eta}$ (dotted) and confidence intervals (10%, dash-dotted) on the y-axis and m on the x-axis, for lower TDC of Clayton ($\theta = 1.3$ and $\theta = 2$) and upper TDC of Gumbel ($\theta = 1.3$) copula; from top to bottom.

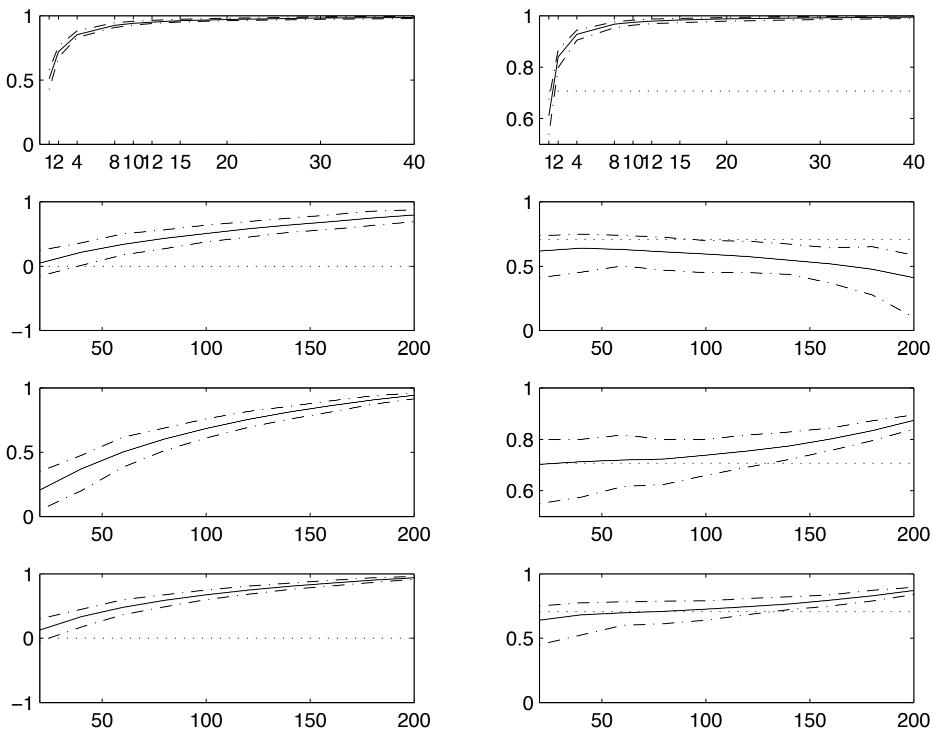

Figure 4.3: Clayton copula with $\theta = 2$, estimators are $\hat{\lambda}^{(4)}, \hat{\lambda}^{(2)}, \hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(3)}$ from top to bottom and for the upper tail dependence left and lower right

For the Gaussian distribution, $\eta < 1$, for all the others $\eta = 1$. Therefore, choosing $m = 15$ seems to be reasonable: for a Gaussian distribution with $\rho = 0.8$, we can reject tail independence but for lower correlation we are able to accept it. For all other distributions with tail dependence, we can reject tail independence as well. This choice of m is consistent with Draisma et al. [2004] who use $m \in \{40, 80, 120\}$ for a sample size of 1,000 observations.

4.2 Estimation of the TDC

Now let us turn to the estimation of the tail dependence coefficient. Figures 4.3, 4.4, 4.5 and 4.6 present the simulation results for the different estimators presented in chapter 3. All figures contain the estimators $\hat{\lambda}^{(4)}, \hat{\lambda}^{(2)}, \hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(3)}$ from top to bottom and for upper tail dependence left, for lower right. Recall that $\hat{\lambda}^{(4)}$ is an estimator in a block-maxima model, therefore, the x-axis refers to the number of blocks ($m = 1$ stands for the original data set, $m = 2$ for a data-set where maxima are calculated for pairs, etc.). For $\hat{\lambda}^{(1)}, \hat{\lambda}^{(2)}$ and $\hat{\lambda}^{(3)}$, the x-axis gives the threshold k. For all figures, the full line is the mean of the estimation results over all 1,000 simulations, the dash-dotted lines are the 2.5%- and the 97.5%- quantiles, the true parameter is dotted.

Figure 4.3 gives the results for a Clayton copula with parameter $\theta = 2$. Recall that in this case, the coefficient of lower tail dependence equals $2^{-1/2} = 0.58$, the upper TDC is zero. $\hat{\lambda}^{(4)}$ performs best for $m = 1$ or $m = 2$, the other estimators give more or less good results when the threshold is smaller than 50.

Figure 4.4 presents the results for the Gumbel copula with parameter $\theta = 2$.

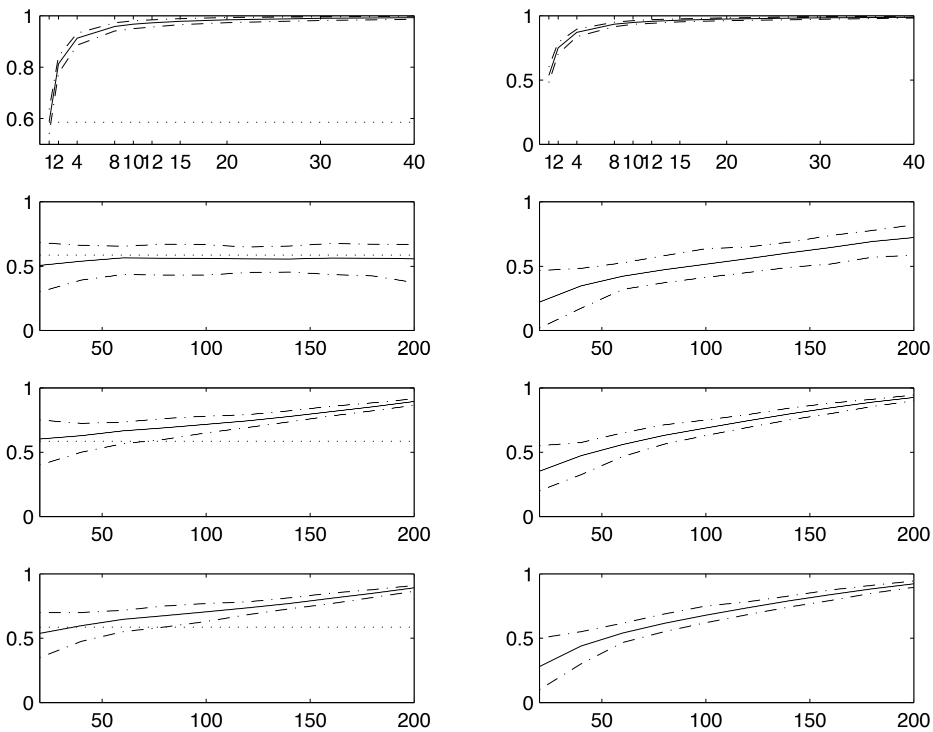

Figure 4.4: Gumbel copula with $\theta = 2$, estimators are $\hat{\lambda}^{(4)}, \hat{\lambda}^{(2)}, \hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(3)}$ from top to bottom and for the upper tail dependence left and lower right

Here, the lower TDC is zero and the upper is equal to $2 - 2^{1/2} = 0.71$. It is not surprising that the block-maxima estimator $\hat{\lambda}_U^{(4)}$ performs best (for $m = 1$ or $m = 2$), since the Gumbel distribution is an extreme value distribution. The other estimators underestimate the upper TDC and overestimate the lower TDC. This again shows that testing for tail independence is important since otherwise, all estimators are biased.

This becomes even clearer when we have a look at the simulation results for the Gaussian distribution in figure 4.5. All estimators overestimate the TDC by far. None of the estimators is near to the true value 0. For the t-distribution with 2 degrees of freedom and $\rho = 0.9$, the TDC is approximately 0.75. Here again, choosing a threshold greater than 50 is not appropriate, as shown in figure 4.6.

Given these simulations, the block-maxima method is done for the original data set, i.e. no block-maxima are calculated, the other estimators are calculated for $k = 50$ since this seems to be overall a good compromise. Furthermore, as the graphs suggest, $\hat{\lambda}^{(2)}$ and $\hat{\lambda}^{(3)}$ give comparable results (which is plausible when looking at their definition). Therefore, in the following, only $\hat{\lambda}^{(2)}$ will be used.

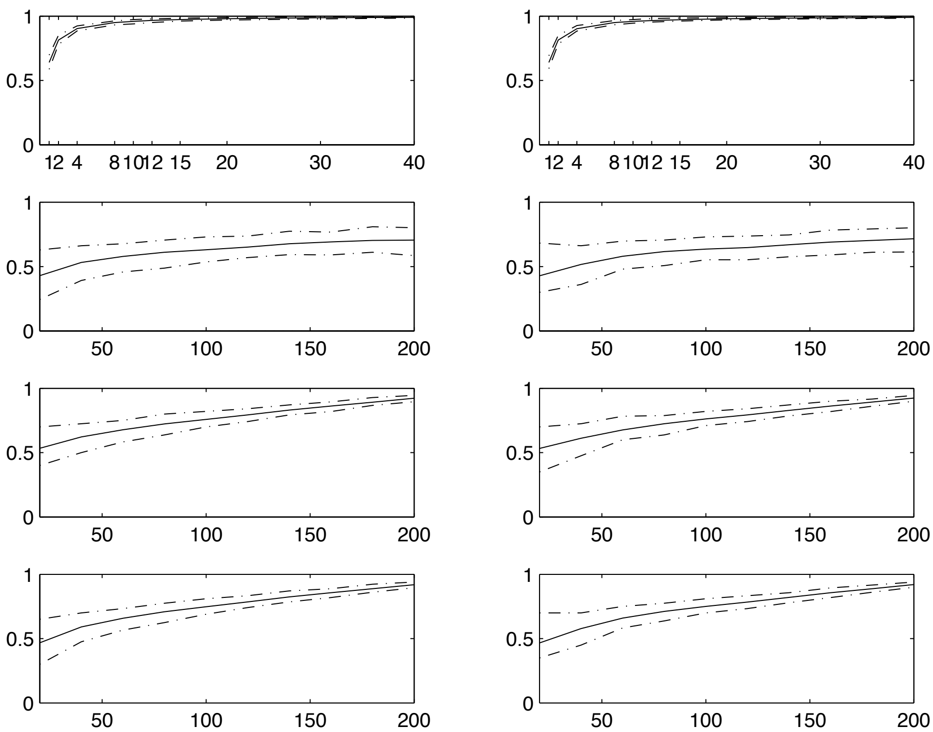

Figure 4.5: Gaussian distribution with $\rho = 0.8$, estimators are $\hat{\lambda}^{(4)}, \hat{\lambda}^{(2)}, \hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(3)}$ from top to bottom and for the upper tail dependence left and lower right

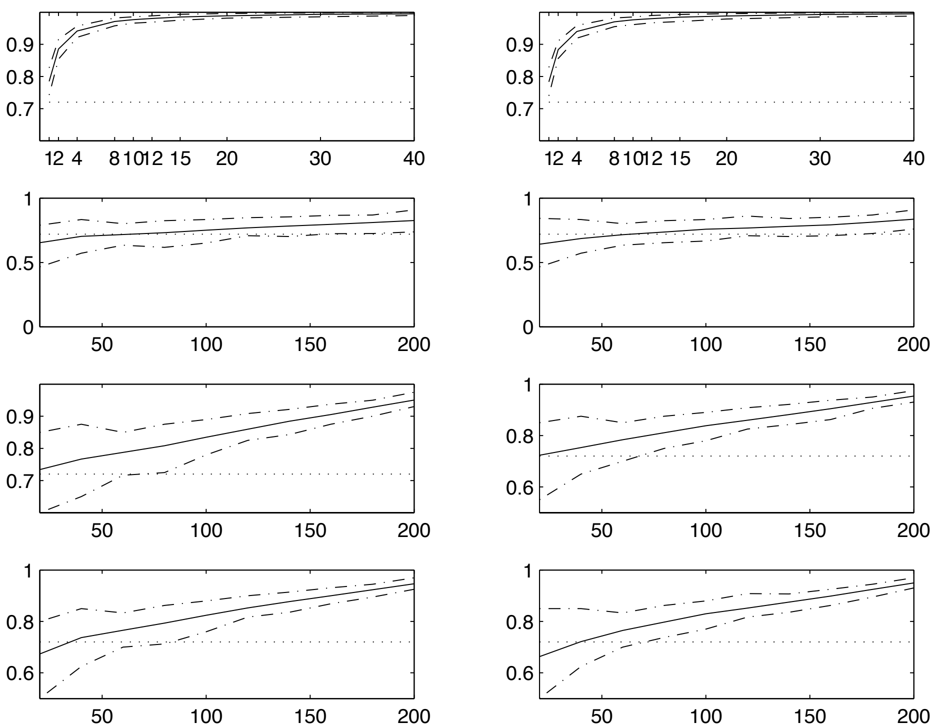

Figure 4.6: t-distribution with $\rho = 0.9, \nu = 2$, estimators are $\hat{\lambda}^{(4)}, \hat{\lambda}^{(2)}, \hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(3)}$ from top to bottom and for the upper tail dependence left and lower right

Chapter 5

Empirical analysis

Given the results of the simulation, we can now turn to the empirical analysis. First of all, the data is presented. Then, the way the data is transformed via a GARCH(1,1) is described. Finally, the estimation of the residual dependence index is carried out and the results of the different estimators of the tail dependence coefficient are given.

5.1 The data

Nine different data sets are analyzed, which cover different sectors: shares of high technology enterprises, insurances and automobile producers, as well as currencies and stock market indices. The analysis is carries out using day-to-day log-returns and a sliding window with window length of 250 data points. Below, the beginning date is given, as well as the date where the estimation results begin (in brackets), i.e. the beginning date plus 250 days. The number of data points given corresponds to the number of points represented in the graphs.

- D_1: Apple and Balda, from 10/02/2000 (09/17/2001) to 08/24/2007 (1550 data points)
- D_2: Balda and Nokia, from 12/27/1999 (12/11/2000) to 08/24/2007 (1750 data points)
- D_3: Cisco and Microsoft, from 01/11/1999 (12/27/1999) to 08/24/2007 (2000 data points)
- D_4: Intel and Microsoft, from 01/11/1999 (12/27/1999) to 08/24/2007 (2000 data points)
- D_5: Münchener Rück and Hannover Rück, from 01/11/1999 (12/27/1999) to 08/24/2007 (2000 data points)
- D_6: Forint and Zloty, from 01/11/1999 (12/27/1999) to 08/24/2007 (2000 data points)
- D_7: Porsche and VW, from 01/11/1999 (12/27/1999) to 08/24/2007 (2000 data points)

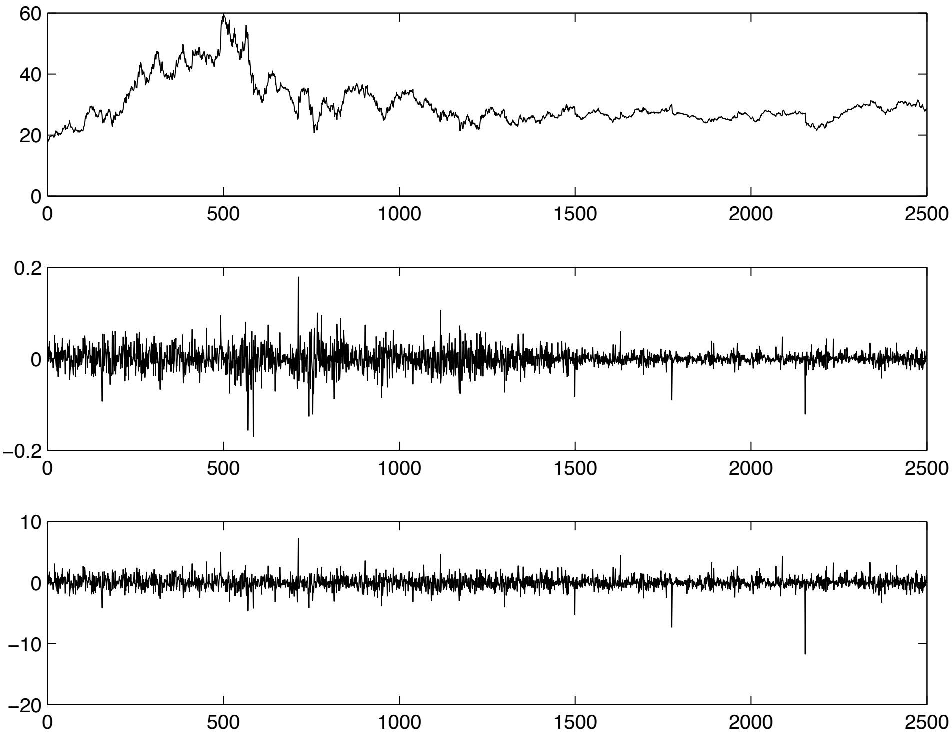

Figure 5.1: Microsoft data for the period 01/26/1998 to 08/24/2007 (2500 data points). Above: stock price, in the middle: day-to-day log-returns; below: estimated innovations of the GARCH(1,1).

- D_8: Allianz and Münchener Rück, from 01/11/1999 (12/27/1999) to 08/24/2007 (2000 data points)
- D_9: Dax and FTSE, from 01/11/1999 (12/27/1999) to 08/24/2007 (2000 data points)

5.2 Modeling the process of the log-returns

Since the pioneer work of Engle [1982] and Bollerslev [1986], ARCH and GARCH (generalized autoregressive conditional heteroscedasticity) are recognized to model well financial markets and their use has become widespread.

The theory presented above is valid only in case of independent variables. Since returns cannot be assumed to be independent, a model has to be specified. Here, a GARCH(1,1) is assumed (see e.g. Chen and Fan [2006]).

Assume that the day-to-day log-returns $\{Y_{jt}\}$ follow a GARCH(1,1), i.e.:

$$\begin{aligned}
Y_{jt} &= \mu_{j,t} + \sqrt{h_{jt}}\varepsilon_{jt} \\
h_{jt} &= \kappa_j + \beta_j h_{j,t-1} + \gamma_j (Y_{j,t-1} - \mu_{j,t-1})^2,
\end{aligned}$$

where $\kappa_j > 0, \beta_j \geq 0, \gamma_j \geq 0$ and $\beta_j + \gamma_j < 1$.

In the following analysis, the empirical residuals $\{\hat{\varepsilon}_{jt}\}$ are used to carry out the different estimations. Figure 5.1 shows an example of how the data is transformed for day-to-day log-returns of the Microsoft stock. It can be seen that in the middle, the variation is especially high during the dot-com boom, then decreases until 2005 and increases again since. After transformation by a GARCH(1,1), the variance seems to be more or less constant, as it is typical for a GARCH(1,1) in a financial context.

5.3 Estimation of the TDC and testing for tail independence

The following figures give the estimation results for the different data sets D_1 through D_9. For each data set, there are two figures: the estimated residual dependence index $\hat{\eta}$ in order to test for tail independence and the estimated TDCs. The upper part of each figure depicts the estimation of $\hat{\eta}$ and $\hat{\lambda}$ in the upper tail, the part below in the lower tail. For representational convenience, $\hat{\eta}$ for the lower tail is multiplied by -1 to lie in $[-1, 0]$.

High-technology shares

Figure 5.2 shows the results for the two stocks of Apple and Balda. Since Balda is a supplier for Apple, one could expect the two shares to be strongly linked in a way that tail dependence can be observed. Bad news for Apple should mean bad news for Balda and the launch of a new device should boost both share prices. Interestingly this is not observed, i.e. tail independence cannot be rejected for the whole period analyzed. There is not a single window in which tail independence can be rejected since the upper bound of the confidence interval of $\hat{\eta}$ is always below 1. Furthermore, all estimators of the TDC show low values, $\hat{\lambda}^{(2)}$ even becomes negative. For the shares of Nokia and Balda (D_2, figure 5.4), we have upper tail dependence in year 2001 and lower tail dependence in 2005 as suggested by the test for tail independence. In these periods, estimates for the TDC range between 0.3 and 0.5. The data set of Microsoft and Cisco (figure 5.6) exhibits a stronger dependence: in 2003 and 2006, we can often accept upper tail dependence between both shares. Interestingly, the TDC estimates are lower ($\in [0.1, 0.4]$) for the period of 2006 than in 2003 ($\in [0.4, 0.6]$) even though the values of $\hat{\eta}$ are higher in 2006. Lower tail dependence occurs less often and can be found in 2004. This is as well true for D_4 (Microsoft and Intel, figure 5.8), which in contrast to D_3 has upper tail dependence in 2003.

Insurances

For D_5 (Münchener Rück and Hannover Rück, see figure 5.10), we find that the link in the upper tail is high in the second semester of 2004, whereas lower tail independence can never be rejected. Nevertheless, all estimates of the TDC (lower and upper) show an increase since 2002, which continues until 2007. Dataset D_8 comprehends Allianz and Münchener Rück and shows strong dependence in the lower tail for the period of 2000 - 2002 and as well in 2006. Upper tail independence can also be rejected for many windows, for details see figure 5.16.

Automobile industry

For D_7 (Porsche and VW, see figure 5.14), there is lower tail dependence in the end of 2002 and upper tail dependence in 2004, according to the tail independence test. Afterwards, neither upper nor lower tail dependence can be accepted. Nonetheless, the estimates of the TDC stay on the same level.

Currencies and stock indices

D_6 and D_9 analyze currencies (Forint and Zloty in D_6, see figure 5.12) and two stock market indices (DAX and FTSE, see figure 5.18). For D_6, we observe strong lower tail dependence since 2006, whereas there is only one shorter period where upper tail dependence exists, namely in the first semester 2004. The two stock market indices are strongly linked, as well: Both, upper and lower tail dependence, can be accepted for a large number of windows, namely in 2002 and between 2005 and 2006.

The empirical analysis reveals that, when using a test for tail independence, there are few data sets, where tail dependence can be found for a long period of time. Interestingly, in times, where tail independence can be rejected, the estimates for the TDC are not necessarily higher than in ones where this is not the case. This again emphasizes the importance of the test for tail independence.

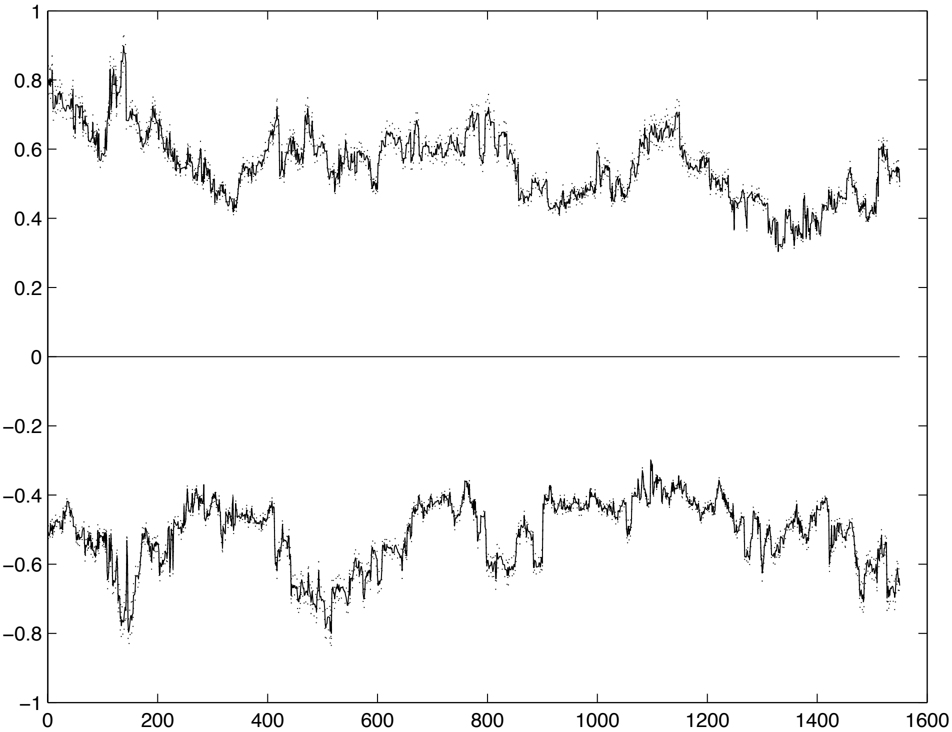

Figure 5.2: Estimation results of D_1, Apple and Balda. Estimations of $\hat{\eta}$ for upper (> 0) and lower (< 0) TDC with respective 10%-confidence intervals.

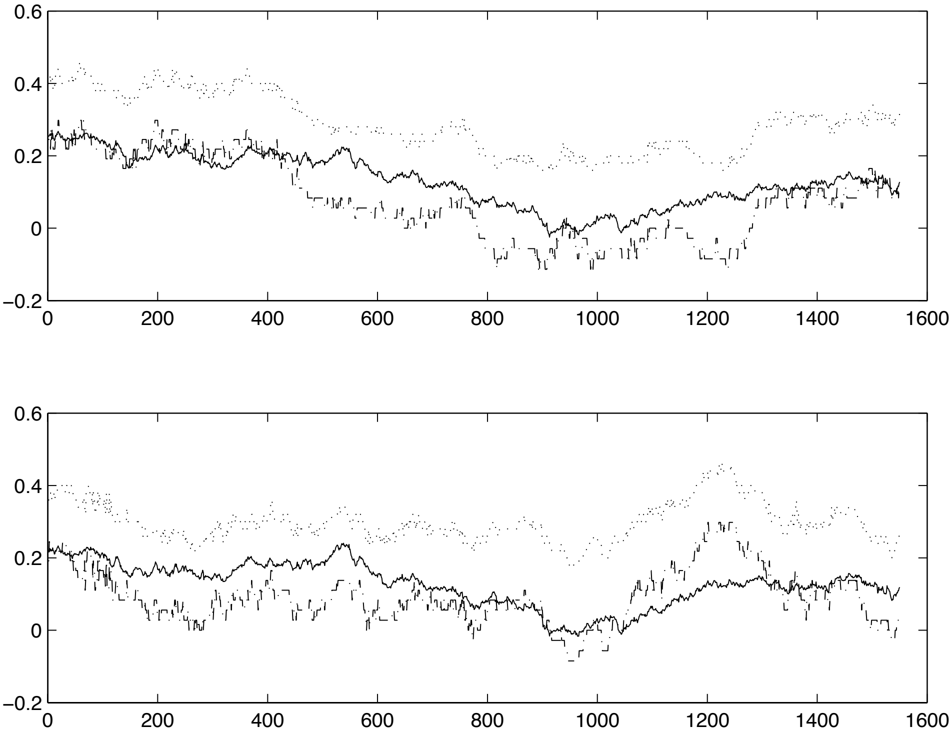

Figure 5.3: Estimation results of D_1, Apple and Balda. Given are $\hat{\lambda}^{(1)}$ dotted, $\hat{\lambda}^{(2)}$ dash-dotted and $\hat{\lambda}^{(4)}$ full line (above for upper, below for lower TDC).

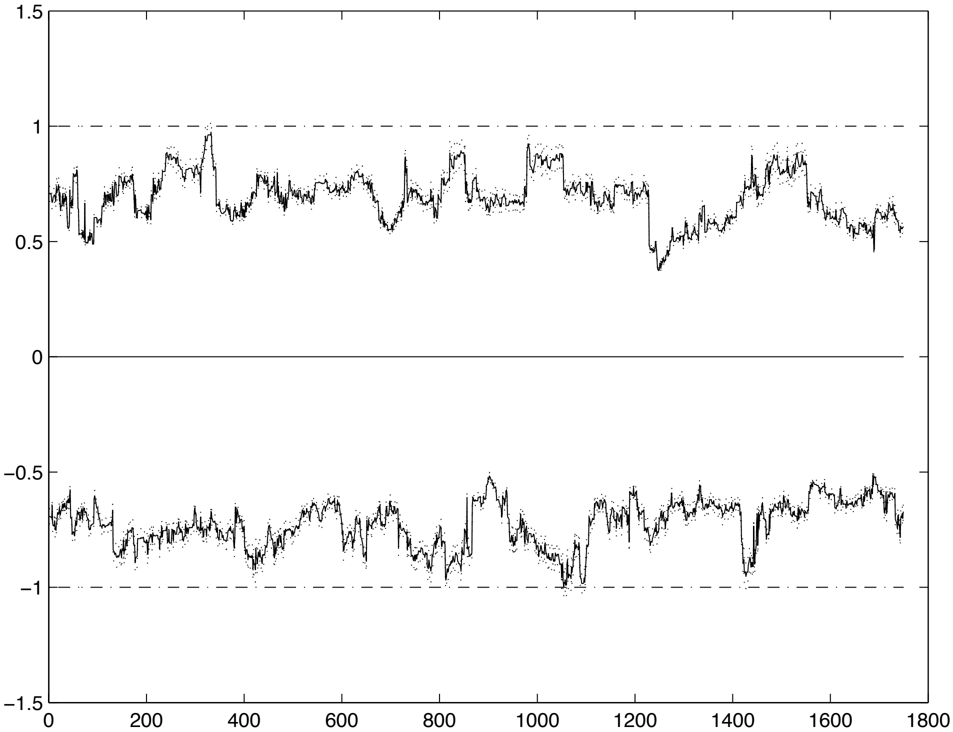

Figure 5.4: Estimation results of D_2, Balda and Nokia. Estimations of $\hat{\eta}$ for upper (> 0) and lower (< 0) TDC with respective 10%-confidence intervals.

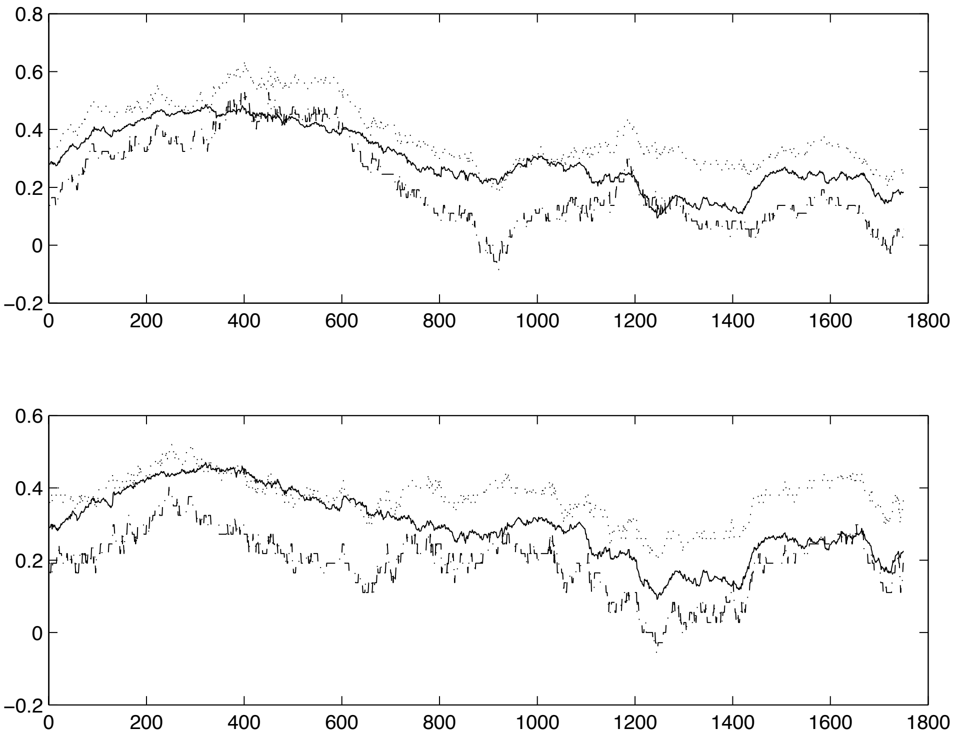

Figure 5.5: Estimation results of D_2, Balda and Nokia. Given are $\hat{\lambda}_U^{(1)}$ dotted, $\hat{\lambda}_U^{(2)}$ dash-dotted and $\hat{\lambda}_U^{(4)}$ full line (above) and $\hat{\lambda}_L^{(1)}$ dotted, $\hat{\lambda}_L^{(2)}$ dash-dotted and $\hat{\lambda}_L^{(4)}$ full line (below).

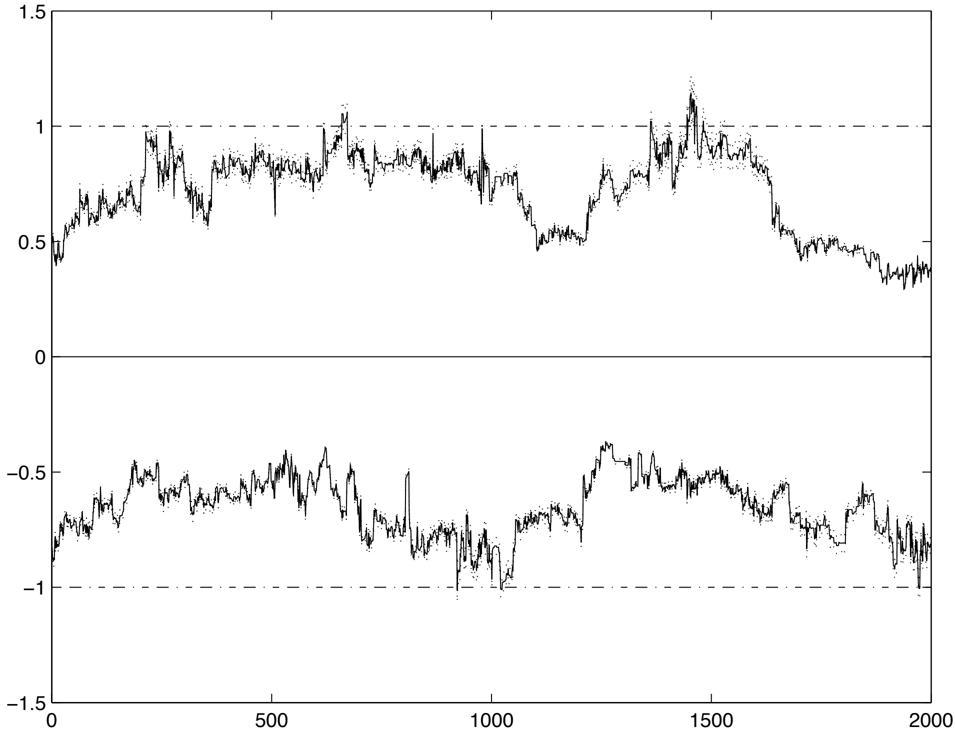

Figure 5.6: Estimation results of D_3, Cisco and Microsoft. Estimations of $\hat{\eta}$ for upper (> 0) and lower (< 0) TDC with respective 10%-confidence intervals.

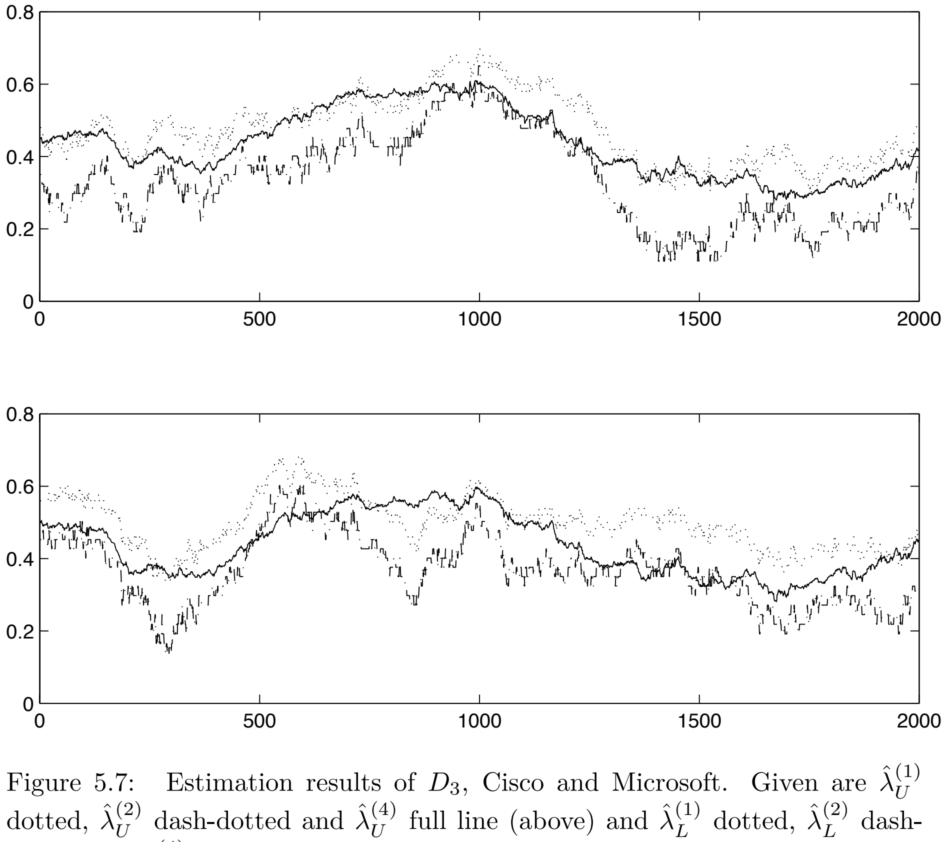

Figure 5.7: Estimation results of D_3, Cisco and Microsoft. Given are $\hat{\lambda}_U^{(1)}$ dotted, $\hat{\lambda}_U^{(2)}$ dash-dotted and $\hat{\lambda}_U^{(4)}$ full line (above) and $\hat{\lambda}_L^{(1)}$ dotted, $\hat{\lambda}_L^{(2)}$ dash-dotted and $\hat{\lambda}_L^{(4)}$ full line (below).

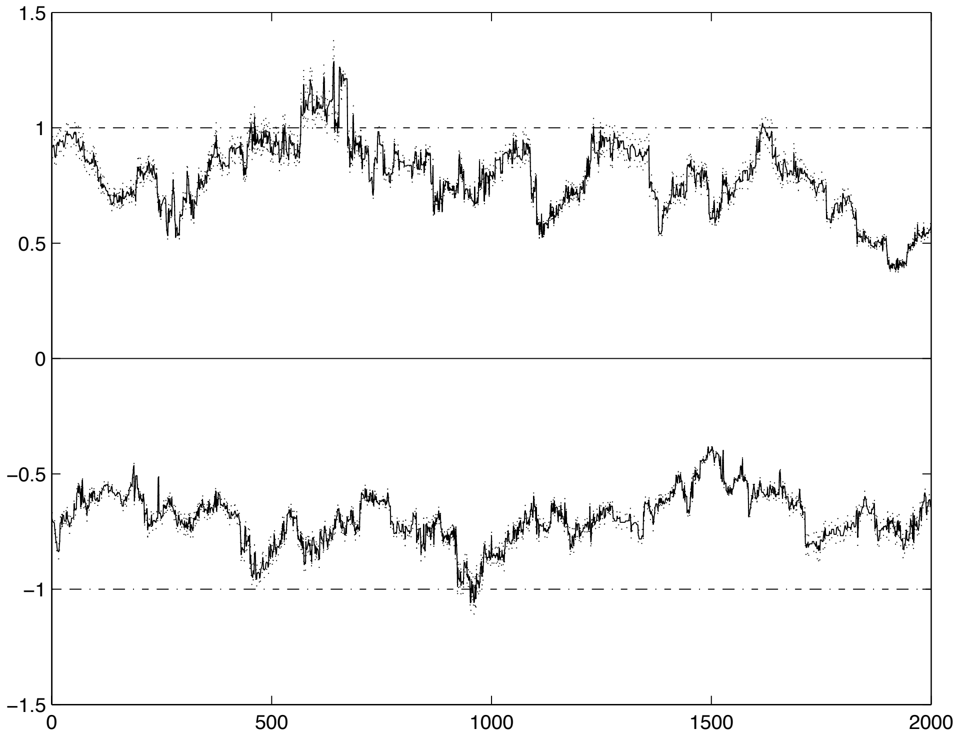

Figure 5.8: Estimation results of D_4, Intel and Microsoft. Estimations of $\hat{\eta}$ for upper (> 0) and lower (< 0) TDC with respective 10%-confidence intervals.

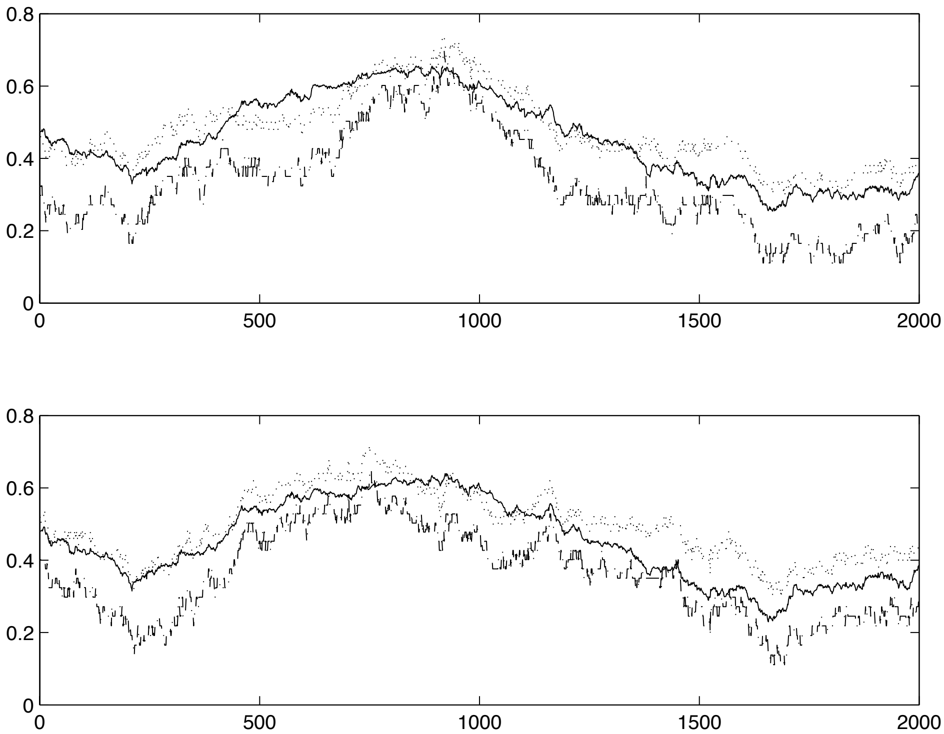

Figure 5.9: Estimation results of D_4, Intel and Microsoft. Given are $\hat{\lambda}_U^{(1)}$ dotted, $\hat{\lambda}_U^{(2)}$ dash-dotted and $\hat{\lambda}_U^{(4)}$ full line (above) and $\hat{\lambda}_L^{(1)}$ dotted, $\hat{\lambda}_L^{(2)}$ dash-dotted and $\hat{\lambda}_L^{(4)}$ full line (below).

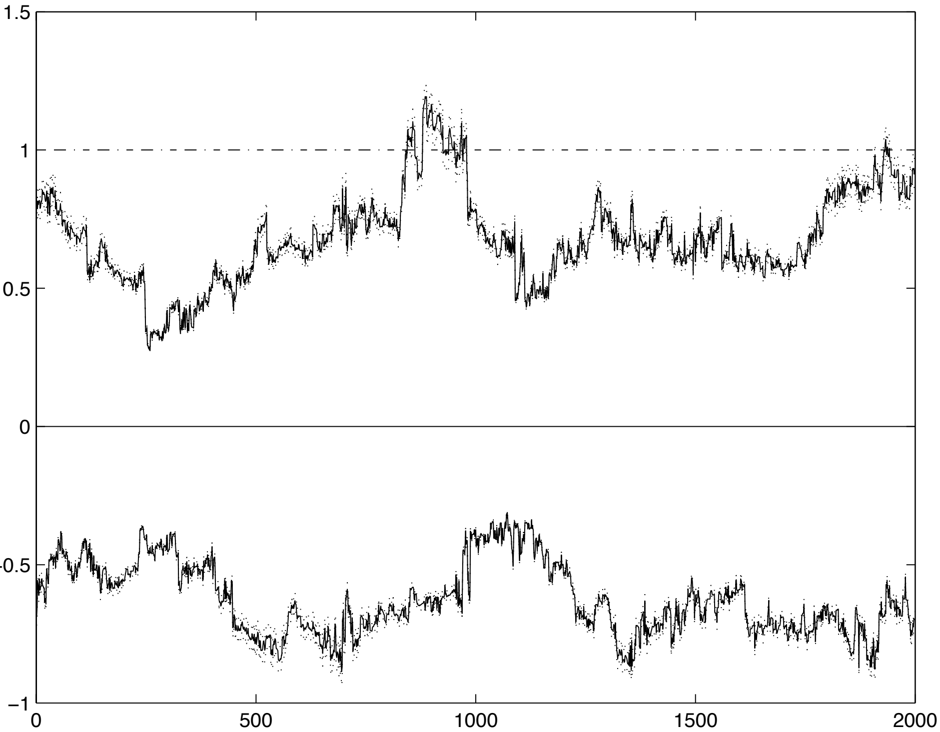

Figure 5.10: Estimation results of D_5, Münchener Rück and Hannover Rück. Estimations of $\hat{\eta}$ for upper (> 0) and lower (< 0) TDC with respective 10%-confidence intervals.

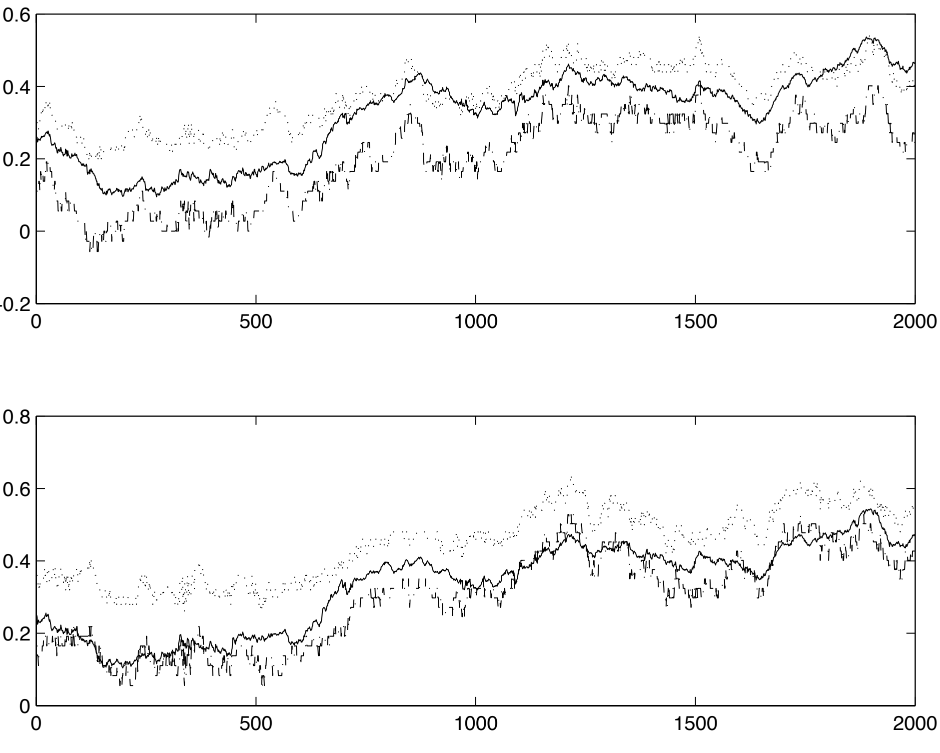

Figure 5.11: Estimation results of D_5, Münchener Rück and Hannover Rück. Given are $\hat{\lambda}_U^{(1)}$ dotted, $\hat{\lambda}_U^{(2)}$ dash-dotted and $\hat{\lambda}_U^{(4)}$ full line (above) and $\hat{\lambda}_L^{(1)}$ dotted, $\hat{\lambda}_L^{(2)}$ dash-dotted and $\hat{\lambda}_L^{(4)}$ full line (below).

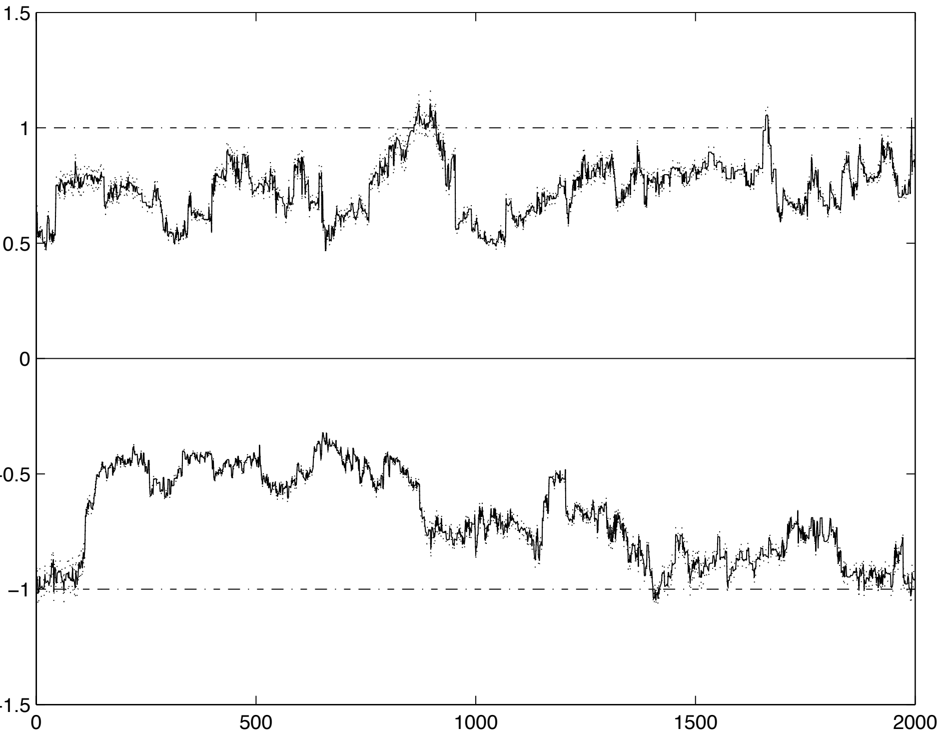

Figure 5.12: Estimation results of D_6, Forint and Zloty. Estimations of $\hat{\eta}$ for upper (> 0) and lower (< 0) TDC with respective 10%-confidence intervals.

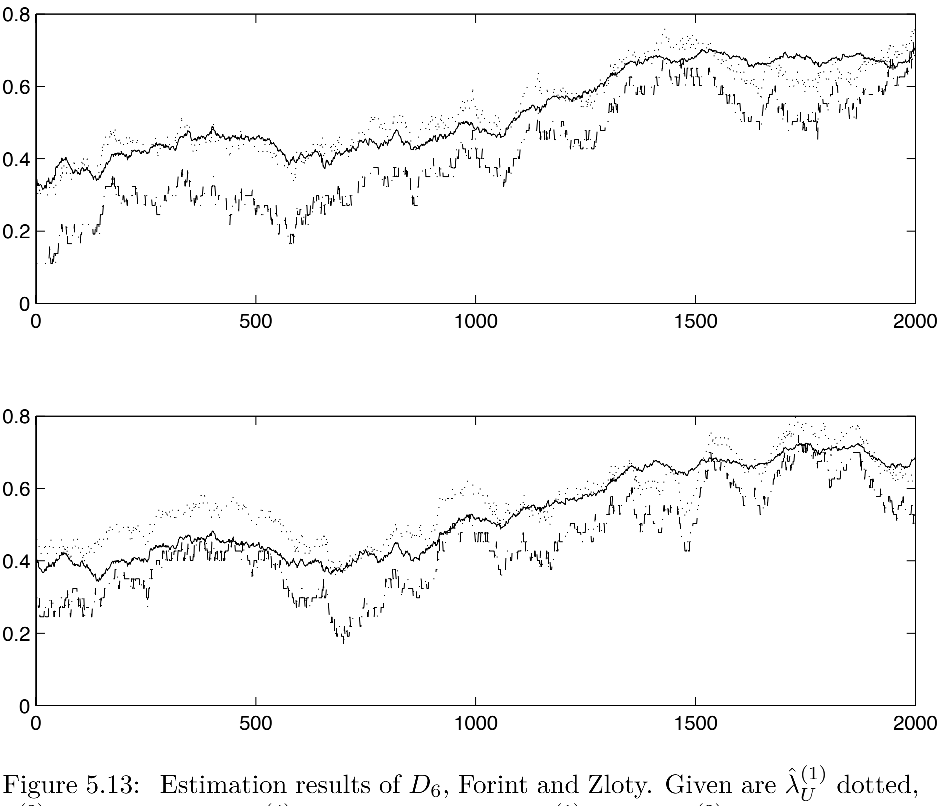

Figure 5.13: Estimation results of D_6, Forint and Zloty. Given are $\hat{\lambda}_U^{(1)}$ dotted, $\hat{\lambda}_U^{(2)}$ dash-dotted and $\hat{\lambda}_U^{(4)}$ full line (above) and $\hat{\lambda}_L^{(1)}$ dotted, $\hat{\lambda}_L^{(2)}$ dash-dotted and $\hat{\lambda}_L^{(4)}$ full line (below).

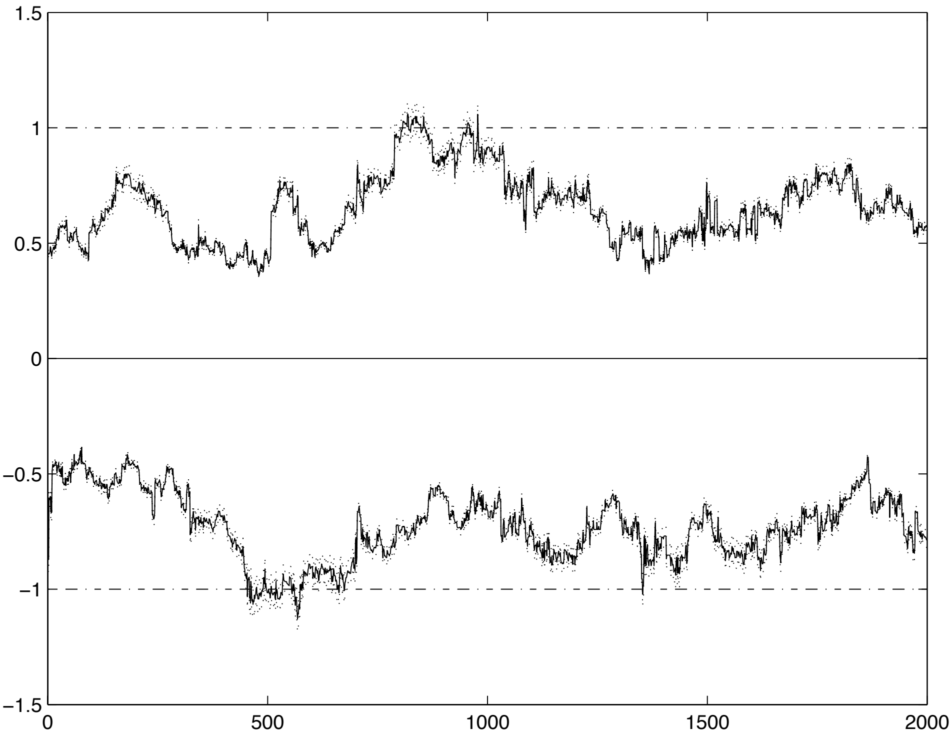

Figure 5.14: Estimation results of D_7, Porsche and VW. Estimations of $\hat{\eta}$ for upper (> 0) and lower (< 0) TDC with respective 10%-confidence intervals.

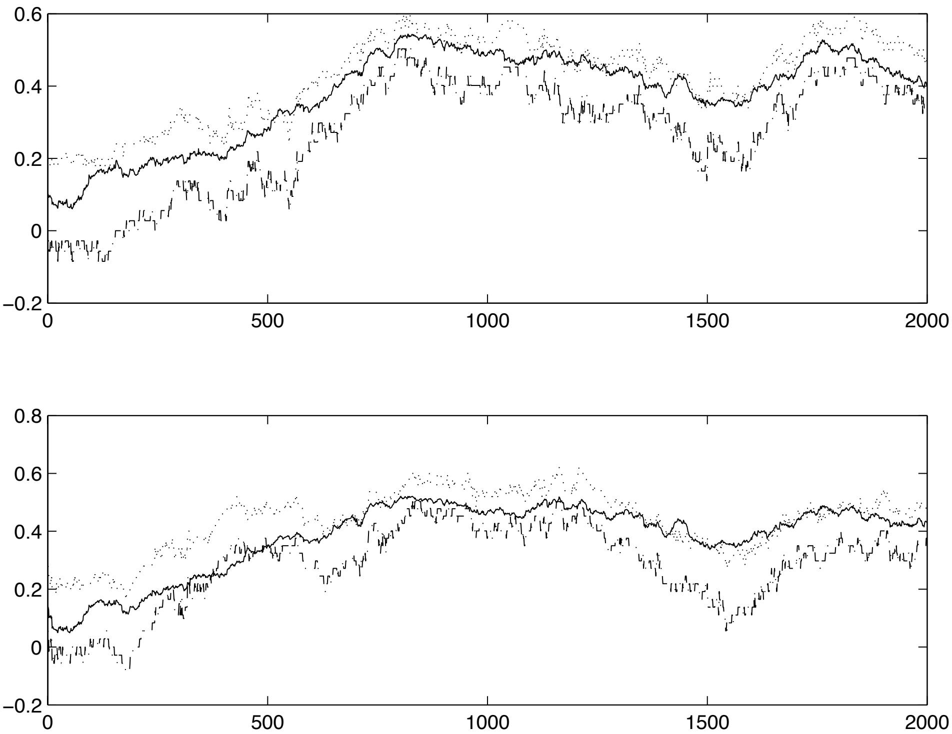

Figure 5.15: Estimation results of D_7, Porsche and VW. Given are $\hat{\lambda}_U^{(1)}$ dotted, $\hat{\lambda}_U^{(2)}$ dash-dotted and $\hat{\lambda}_U^{(4)}$ full line (above) and $\hat{\lambda}_L^{(1)}$ dotted, $\hat{\lambda}_L^{(2)}$ dash-dotted and $\hat{\lambda}_L^{(4)}$ full line (below).

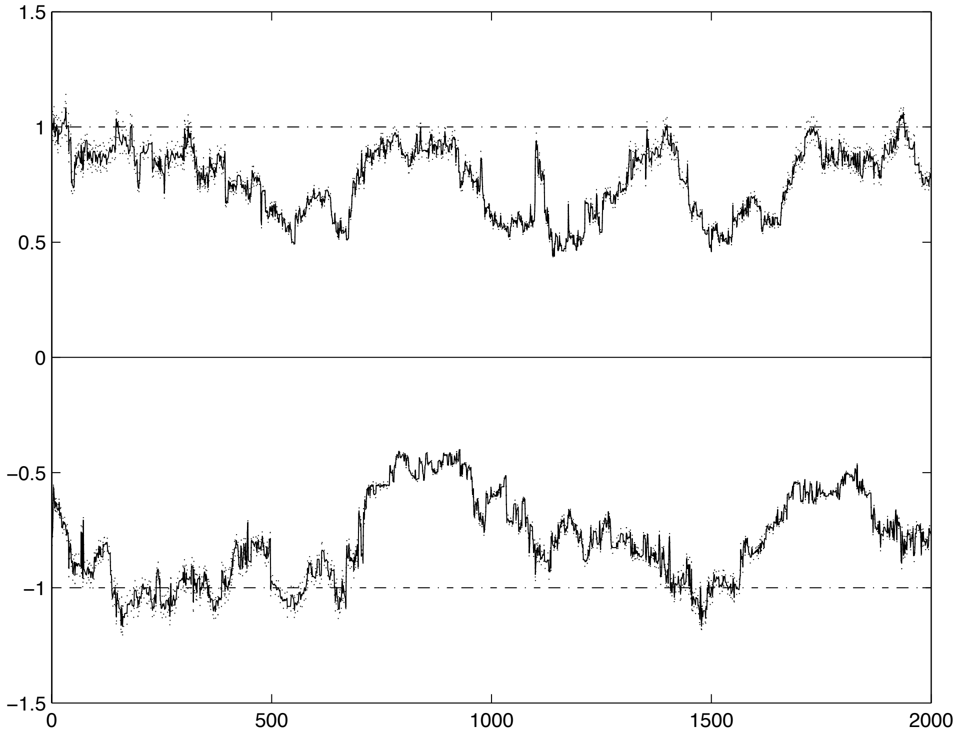

Figure 5.16: Estimation results of D_8, Allianz and Münchener Rück. Estimations of $\hat{\eta}$ for upper (> 0) and lower (< 0) TDC with respective 10%-confidence intervals.

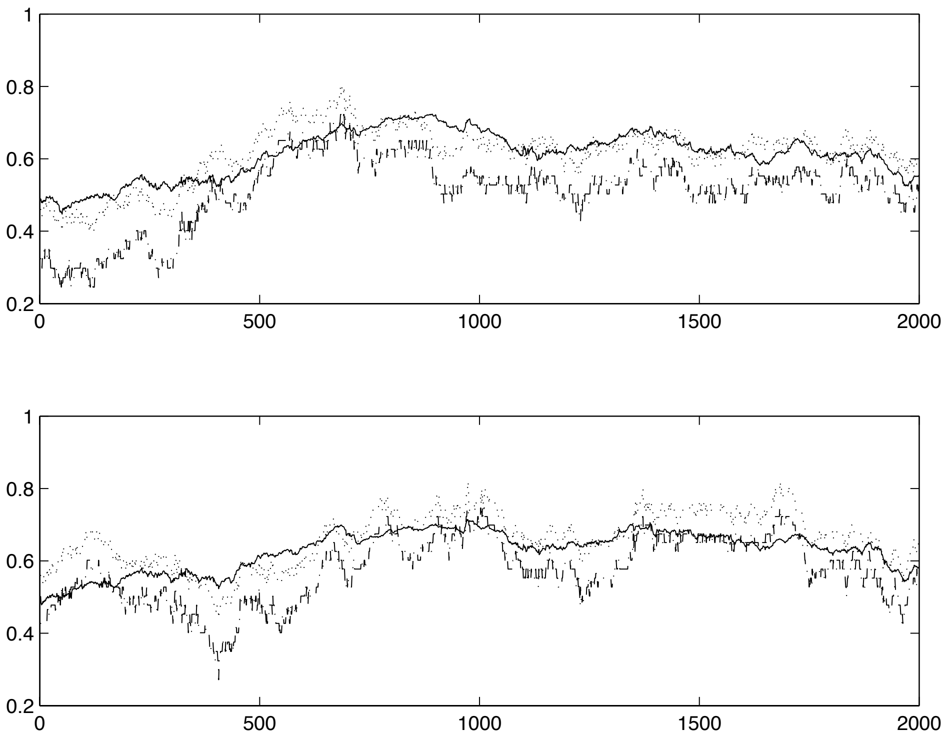

Figure 5.17: Estimation results of D_8, Allianz and Münchener Rück. Given are $\hat{\lambda}_U^{(1)}$ dotted, $\hat{\lambda}_U^{(2)}$ dash-dotted and $\hat{\lambda}_U^{(4)}$ full line (above) and $\hat{\lambda}_L^{(1)}$ dotted, $\hat{\lambda}_L^{(2)}$ dash-dotted and $\hat{\lambda}_L^{(4)}$ full line (below).

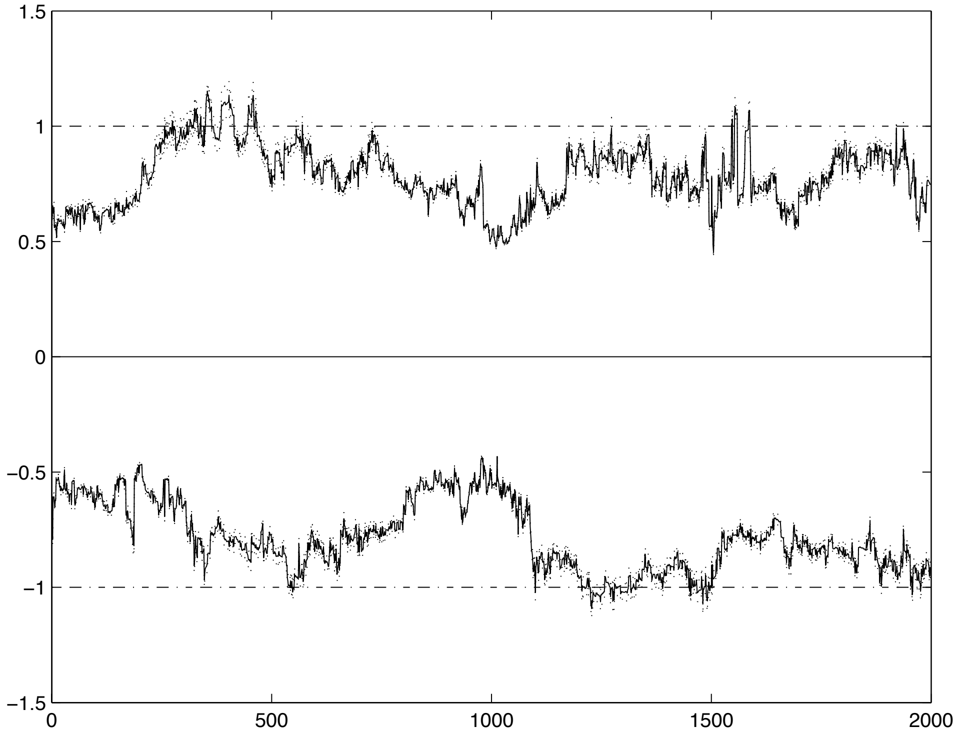

Figure 5.18: Estimation results of D_9, Dax and FTSE. Estimations of $\hat{\eta}$ for upper (> 0) and lower (< 0) TDC with respective 10%-confidence intervals.

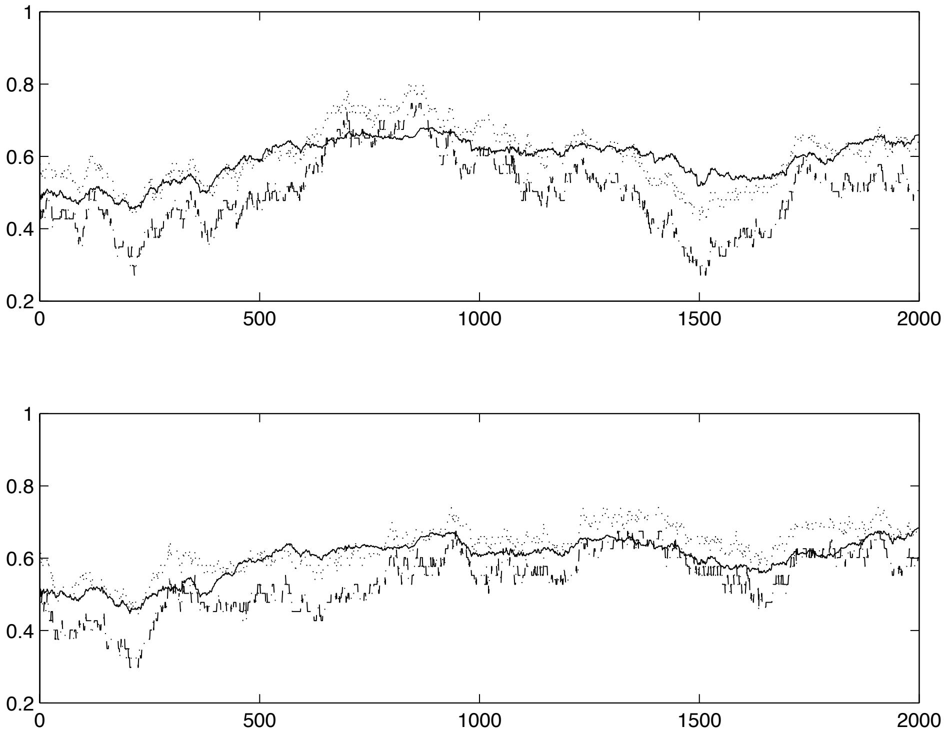

Figure 5.19: Estimation results of D_9, Dax and FTSE. Given are $\hat{\lambda}_U^{(1)}$ dotted, $\hat{\lambda}_U^{(2)}$ dash-dotted and $\hat{\lambda}_U^{(4)}$ full line (above) and $\hat{\lambda}_L^{(1)}$ dotted, $\hat{\lambda}_L^{(2)}$ dash-dotted and $\hat{\lambda}_L^{(4)}$ full line (below).

Chapter 6

Conclusion

The aim of this thesis is to enlighten two concepts that are en vogue at the moment, as well in applied statistics and finance as for practitioners in the financial sector: copulae as a means to separate marginal distributions from dependence and tail dependence to quantify the probability of joint events in the tail of a distribution. As seen, copulae can be used to model tail dependence. An important theoretical concept used throughout this thesis is bivariate Extreme Value Theory (EVT), which tries precisely to analyze the behavior at the tail of a bivariate distribution.

Since estimating the probability of something that happens rarely is hard, the techniques are not always very precise. But as J. Tawn puts it (quoted by Embrechts et al. [2003] in their preface): "My answer to sceptics is that if people aren't given well-founded methods like EVT, they'll use dubious ones instead". Therefore, even if results can be sometimes unsatisfactory, it is hard to do better, since predicting the rare events is a tough job. And EVT does a lot better than other methods. Especially in finance, where the quantification of extreme events plays an important role, the use of EVT could make different concepts more objective. Unfortunately, these techniques are not yet prevalent, perhaps due to the theoretical complexity, which makes EVT mainly a subject of mathematical statistics and the fact that even in mathematics, multivariate EVT is a relatively new and active field of research.

One important feature of this thesis is the implementation of the test for tail independence, which is recognized to be indispensable but rarely utilized in a financial context. This test permits to detect the periods where tail dependence exists. Afterwards, the TDC can be estimated using the four estimators presented in this thesis. Omitting the test for tail independence would introduce a large bias in the estimation and make it difficult to decide whether there is just correlation or in fact tail dependence. This is in fact often done in the literature and therefore some authors come to the conclusion that "Tail dependence is indeed often found in financial data series" (Schmidt [2005], p. 83). As presented in this thesis, the phenomenon is less common, the periods where indeed tail independence can be rejected are few. But that makes it in a way even more appealing: in the future, more research could be precisely done to investigate the reason why in these periods some assets are strongly linked and what this means for risk management. Another approach would be to compare the periods of tail dependence to some macro-economic variables, e.g. in a time series model. Nevertheless, working with rare events stays complex and even testing for tail independence does not overcome all problems.

Future research could also try to make the estimation procedure adaptive, i.e. to estimate the parameters using a locally optimal window length instead of a fixed window length of 250. But therefore, it would be helpful to test the assumptions of a GEV or a GPD and to see if in a small sample like 250 observations, one can reasonably accept the distributional assumptions. A different direction for future research would be to compare the different test of tail independence of sections 3.7.1 and 3.7.2, especially in their small sample behavior. Overall, it would be interesting to integrate the concept of tail dependence in a larger framework, i.e. to use the information provided by TDC estimates e.g. in risk management.

Bibliography

T. Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

P. Capéraà, A. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577, 1997.

X. Chen and Y. Fan. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135:125–154, 2006.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, springer series in statistics edition, 2001.

S. Coles, J. Heffernan, and J. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2:339–365, 1999.

A. Dias and P. Embrechts. Dynamic copula models for multivariate high-frequency data in finance, 2004.

G. Draisma, H. Drees, A. Ferreira, and L. de Haan. Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10(2):251–280, 2004.

P. Embrechts, F. Lindskog, and A. McNeil. Modelling Dependence with copulas and Applications to Risk Management. In S. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, chapter 8, pages 329–384. Elsevier, 2001.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, Applications of Mathematics, Stochastic Modelling and Applied Probability 33 edition, 2003.

R. Engle. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation. *Econometrica*, 50, 1982.

M. Falk and R. Michel. Testing for tail independence in extreme value models. *Annals of the Institute of Statistical Mathematics*, 58, 2006.

M. Falk and R. Reiss. Efficient estimation of the canonical dependence function. *Extremes*, 6, 2003.

M. Falk and R. Reiss. On Pickands coordinates in arbitrary dimensions. *Journal of Multivariate Analysis*, 92, 2005a.

M. Falk and R. Reiss. On the distribution of Pickands coordinates in bivariate EV and GP models. *Journal of Multivariate Analysis*, 93, 2005b.

V. Fernandez. Risk management under extreme events. *International Review of Financial Analysis*, 14, 2005.

A. Ferreira and L. de Haan. *Extreme Value Theory: An Introduction*. Springer, Springer Series in Operations Research and Financial Engeneering edition, 2006.

G. Frahm, M. Junker, and R. Schmidt. Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 37, 2005.

J. Heffernan. A Directory of Coefficients of Tail Dependence. *Extremes*, 3, 2000.

H. Joe. *Multivariate Models and Dependence Concepts*. Chapman Hall, Monographs on Statistics and Applied probability 73 edition, 1997.

M. Junker and A. May. Measurement of aggregate risk with copulas. *Econometrics Journal*, 8, 2005.

G. Kim, J. Silvapulle, and P. Silvapulle. Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics and Data Analysis*, 51, 2007.

C. Klüppelberg, G. Kuhn, and L. Peng. Estimating Tail dependence of Elliptical Distributions. *Working paper 470 of Sonderforschungsbereich 386*, 2006.

E. Kole, K. Koedijk, and M. Verbeek. Selecting copulas for risk management. *Journal of Banking and Finance*, 31, 2007.

S. Kotz and S. Nadarajah. *Extreme Value Distributions*. Imperial College Press, 2000.

A. Ledford and J. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83, 1996.

A. Ledford and J. Tawn. Modelling dependence within Joint Tail Regions. *Journal of the Royal Statistical Society*, 59 Section B(2), 1997.

A. Ledford and J. Tawn. Concomitant tail behaviour for extremes. *Advances in Applied Probability*, 30, 1998.

T. Mikosch. Copulas: Tales and facts. http://www.math.ku.dk/∼mikosch/maphysto_extremes_2005/s.pdf, 2005.

R. Nelsen. *An introduction to copulas*. Springer, 2006.

L. Peng. Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics and Probability Letters*, 43, 1999.

J. Pickands. Multivariate extreme value distributions. *Bulletin of the International Statistical Institute*, 49, 1981.

J. Rodriguez. Measuring financial contagion: A Copula approach. *Journal of Empirical Finance*, 14, 2007.

R. Schmidt. Tail dependence. In P. Cizek, W. Härdle, and R. Weron, editors, *Statistical Tools in Finance and Insurance*. Springer, 2005.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 1959.